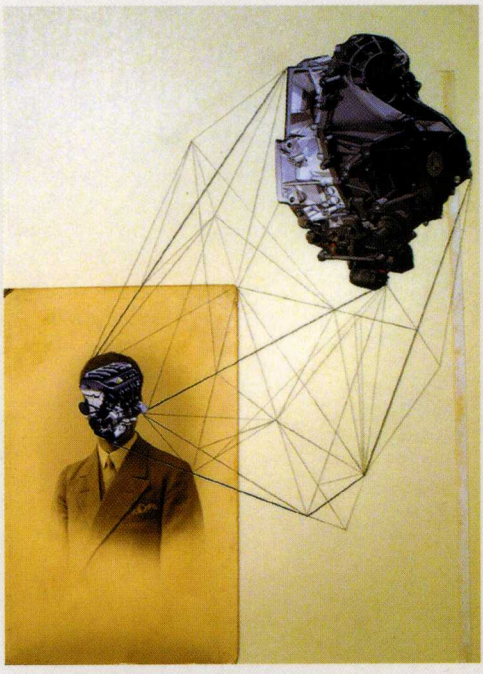


Bibliothèque

Gerald M. Edelman

Biologie de la conscience

Odile Jacob

BIOLOGIE DE LA CONSCIENCE

GERALD M. EDELMAN

BIOLOGIE DE LA CONSCIENCE

Traduit de l'anglais (États-Unis)
par Ana Gerschenfeld

Odile Jacob

Traduit avec le concours du Centre National des Lettres.

L'édition originale en langue anglaise de cet ouvrage est parue sous le titre :
Bright Air, Brilliant Fire : On the Matter of Mind.
© Basic Books, 1992.
Publié en accord avec Basic Books,
Département de HarperCollins Publishers inc., 1992.
Pour la traduction française :
© ODILE JACOB, 2008
15, RUE SOUFFLOT, 75005 PARIS
ISBN 978-2-7381-2071-7

La loi du 11 mars 1957 interdit les copies ou reproductions destinées à une utilisation collective. Toute représentation ou reproduction intégrale ou partielle faite par quelque procédé que ce soit, sans le consentement de l'auteur ou de ses ayants cause, est illicite et constitue une contrefaçon sanctionnée par les articles 425 et suivants du Code pénal.

À la mémoire de deux pionniers intellectuels,
Charles Darwin et Sigmund Freud.
Plus on a de science, plus on a de peine.

C'est par la terre [qui est en nous] que nous connaissons la terre, par l'eau que nous connaissons l'eau, par l'éther, l'éther divin, par le feu, le feu destructeur.

Empédocle

Et, en continuant, nous arrivons à des choses comme le mal, la beauté et l'espoir [...]

« Lequel des deux extrêmes est plus proche de Dieu, si je puis me permettre d'utiliser une métaphore religieuse ? La beauté et l'espoir, ou les lois fondamentales ? Je pense que la bonne démarche consiste, bien sûr, à dire que nous devons considérer l'ensemble des interconnexions structurelles des choses, et que toutes les sciences – et non seulement les sciences, mais tous les efforts de type intellectuel – sont des tentatives visant à découvrir les liaisons des hiérarchies, à relier la beauté à l'histoire, l'histoire à la psychologie humaine, la psychologie humaine au fonctionnement du cerveau, le cerveau aux influx nerveux, les influx nerveux à la chimie, et ainsi de suite, vers le haut et vers le bas, dans un sens comme dans l'autre. Mais aujourd'hui nous ne pouvons pas encore, et cela ne sert à rien de prétendre le contraire, relier exactement et totalement un extrême de cette chose à l'autre, puisque nous ne faisons que commencer à voir que cette hiérarchie relative existe.

Et je ne pense pas que l'un des extrêmes soit plus proche de Dieu que l'autre.

Richard Feynman

Préface

J'ai écrit ce livre parce que je pense que le thème sur lequel il porte est le plus important que l'on puisse imaginer. Nous en sommes au début de la révolution des neurosciences. Lorsqu'elle sera achevée, nous saurons comment fonctionne l'esprit, nous comprendrons ce qui régit notre nature, et aussi comment nous faisons pour connaître le monde. En fait, ce qui se passe actuellement en neurosciences peut être considéré comme le prélude à la plus grande des révolutions scientifiques, une révolution aux répercussions sociales inévitables et fondamentales.

Mais ceci n'est pas un livre scientifique, du moins au sens strict du terme. C'est un livre sur la science et aussi sur mes opinions personnelles. Pour l'écrire et parvenir à expliquer des questions plutôt techniques à des lecteurs non spécialisés, j'ai dû mettre de côté les indispensables précautions oratoires qu'utilisent habituellement entre eux les chercheurs d'une même discipline. Mais tant pis pour les habitudes – surtout si je parviens à intéresser à cette affaire ceux qui n'y travaillent pas, et à les convaincre d'y contribuer et de partager l'excitation que l'on ressent lorsqu'on est sur le point de savoir comment nous faisons pour savoir.

PREMIÈRE PARTIE

Les problèmes

Ceux qui considèrent que seul l'esprit est capable de poser des questions, et qu'il n'a jamais été sérieusement démontré que l'esprit puisse exister en dehors du corps, trouveront qu'il est inutile de justifier l'importance du sujet abordé ici. Dans cette partie du livre, je me propose de présenter au lecteur un certain nombre de notions classiques concernant l'esprit. Je compte également indiquer quel sera ensuite mon propos : décrire une théorie biologique qui rende compte de la façon dont nous en sommes venus à avoir un esprit. Pour ce faire, j'analyserai l'organisation de la matière qui le sous-tend – les neurones, leurs connexions et les structures qu'ils forment.

Chapitre 1

L'esprit

Cogito, ergo sum.
René Descartes

Le défaut du Discours de la Méthode *de Descartes réside dans le parti pris par l'auteur de se vider lui-même de lui-même – de Descartes, de l'homme qu'il était réellement, de l'homme en chair et en os, de l'homme qui ne voulait pas mourir – afin de devenir un simple penseur, c'est-à-dire une abstraction. Mais l'homme qu'il était réellement est revenu et il a fait irruption dans la philosophie de Descartes* [...]

La vérité, c'est sum, ergo cogito – *Je suis, donc je pense, bien que tout ce qui est ne pense pas. La pensée consciente n'est-elle pas, par-dessus tout, conscience d'être ? La pensée pure est-elle possible, sans la conscience de soi, sans la personnalité ?*

Miguel de Unamuno

« Ne pensez pas à un éléphant. »

Bien entendu, vous y avez pensé. Et moi aussi. Mais où se trouve l'éléphant ? Dans votre esprit, certainement pas dans la pièce – du moins pour la plupart des gens qui sont en train de lire ce livre. Mais, pour *ne pas* y penser, il a fallu que vous sachiez de quoi il s'agissait, que vous vous le remémoriez et même, dans certains cas, que vous en évoquiez une image. Surtout, il a fallu que vous compreniez cette langue et ce petit jeu sur les mots.

Voici un autre petit jeu de mots, pour qui comprend l'anglais : *« What is mind ? No matter. What is matter ? Never mind* [1] *»* nous dit la conclusion à laquelle arriva Descartes au cours de ses réflexions sur ce point. Celles-ci ont marqué les débuts de la philosophie moderne et placé l'esprit hors de portée de la recherche scientifique. Pour Descartes, l'esprit serait une substance particulière, située hors

1. Ces phrases admettent deux traductions différentes : 1) « Qu'est-ce que l'esprit ? Ce n'est pas de la matière. Qu'est-ce que la matière ? Elle n'est jamais esprit » et 2) « Qu'est-ce que l'esprit ? C'est sans importance. Qu'est-ce que la matière ? Peu importe. » (N.d.T.)

de l'espace, et non une substance étendue comme la matière. Depuis, cette doctrine dualiste a constamment hanté, si ce n'est tout le monde, du moins un grand nombre de philosophes ainsi que certains théologiens.

Qu'est-ce que cela signifie avoir un esprit, être conscient ? Tout le monde se pose ce genre de questions à un moment ou à un autre,



FIGURE 1-1

René Descartes (1596-1650), un des fondateurs de la philosophie moderne en même temps qu'un grand mathématicien. Le dualisme qu'il épousa oppose encore les penseurs modernes à propos de l'esprit. Le dualisme cartésien ne pourra probablement être écarté que lorsque nous aurons compris la relation existant entre la conscience et le monde physique.

mais jusqu'à récemment, les scientifiques n'osaient pas les aborder du point de vue scientifique. À présent, au contraire, un nouvel ensemble de disciplines est entré en scène : les neurosciences. Nous avons commencé à accumuler à une vitesse vertigineuse des connaissances scientifiques sur le cerveau. Nous sommes ainsi devenus capables de décrire scientifiquement la façon dont nous voyons, dont nous entendons, dont nous sentons. L'objet le plus complexe de l'univers commence à nous livrer ses secrets.

Qu'est-ce qui nous amène à penser que cela nous apprendra quoi que ce soit sur notre esprit ? Ce que nous savons déjà. De même que nous avons compris que la matière naît d'un agencement particulier des choses, nous devrions être capables de comprendre quels sont les agencements particuliers de choses qui donnent naissance à l'esprit. C'est de cela qu'il s'agit dans ce livre : établir un lien entre ce que nous savons sur notre esprit et ce que nous commençons à savoir sur notre cerveau.

Je compte aborder des sujets très divers : le système nerveux, les ordinateurs, la perception, le langage, l'individualité. J'essayerai de montrer comment ils sont liés les uns aux autres et au fait que nous soyons conscients. Plutôt que de décrire la façon dont nous pensons ou raisonnons, je m'intéresserai aux bases de ces activités supérieures. Mon objectif ultime est de montrer qu'il est scientifiquement possible de comprendre l'esprit. J'essayerai de restreindre les détails techniques au minimum, mais je n'hésiterai pas, le moment venu, à démonter les dogmes et les idées reçues que je crois être faux. Par conséquent, certaines parties de ce livre seront consacrées à mettre à l'index des points de vue que je considère comme insoutenables. Je vous promets néanmoins que le bilan global sera positif et constructif. Après tout, ce sujet, tout comme l'obstétrique, est essentiel à notre présence ici. Il est au centre de nos préoccupations.

Commençons donc.

Le mot *esprit* évoque aussitôt d'obscures discussions philosophiques. Mais il projette également une ombre familière, issue de la vie quotidienne : « avoir l'esprit ailleurs », « esprit de famille », « mot d'esprit », etc. Ce à quoi renvoient ces expressions n'est nullement clair. Mais nous pouvons quand même nous fier à notre bon sens pour commencer :

1. Les objets n'ont pas d'esprit.
2. Les êtres humains normaux ont un esprit ; certains animaux se comportent comme s'ils en avaient un.

3. Les êtres qui ont un esprit sont capables de faire référence à d'autres êtres ou à des objets ; les objets dépourvus d'esprit ne font jamais référence à des êtres ou à d'autres objets.

Cette dernière propriété, que le philosophe allemand Franz Brentano appelait intentionnalité, est un bon indicateur de l'existence d'un processus mental. Elle fait référence à l'idée que la conscience est toujours conscience *de* quelque chose, qu'elle a toujours un objet. Je ferai fréquemment appel à cette notion d'intentionnalité dans ce qui suit.

Toutefois, il ne suffit pas de disposer d'un ensemble d'indicateurs pour décrire la relation entre l'esprit et la matière, et en particulier entre l'esprit et l'organisation spécifique de la matière qui le sous-tend. Il n'est pas surprenant que les gens considèrent l'esprit lui-même comme une réalité particulière, ou comme une substance d'un type tout à fait spécial. Après tout, il semble si différent de la matière ordinaire que celui qui en possède un a du mal à conclure, par simple introspection, que l'esprit pourrait être issu d'*interactions* se produisant au sein d'une matière dépourvue d'intentions. Mais, comme l'a fait remarquer William James, l'esprit est un processus, pas une substance. La recherche scientifique moderne indique que des processus extraordinaires sont susceptibles de se faire jour dans la matière ; en fait, la matière elle-même peut être considérée comme le résultat de processus d'échange d'énergie. La science moderne a redéfini la matière en termes de processus ; elle n'a pas redéfini l'esprit comme étant une forme particulière de la matière. Tout au long de ce livre, la thèse fondamentale que je compte défendre consiste à dire que l'esprit est un processus d'un type particulier qui dépend de certaines formes particulières d'organisation de la matière.

Revenons à notre liste de notions intuitives. Nous voyons que les êtres qui semblent posséder un esprit sont des organismes biologiques (et en particulier des animaux). Par conséquent, il est naturel de supposer que c'est un type particulier d'organisation *biologique* qui donne lieu aux processus mentaux. À l'évidence, donc, pour aborder le sujet de manière scientifique, nous devons nous intéresser à la façon dont le cerveau est organisé. Il serait erroné, néanmoins, d'ignorer le reste du corps, parce que les fonctions animales (en particulier le mouvement) et le développement du cerveau sont intimement liés.

Depuis Darwin, chaque fois qu'ils sont confrontés à des formes particulières d'organisation biologique, les biologistes se demandent



FIGURE 1-2

William James (1842-1910), un des fondateurs de la psychologie physiologique moderne en même temps qu'un des principaux représentants de la philosophie pragmatiste. Ses réflexions sur la conscience – l'idée qu'il s'agit d'un processus et non d'une substance, qu'elle est personnelle et qu'elle reflète une intentionnalité – ont profondément marqué notre conception moderne sur le sujet.

presque automatiquement comment l'évolution a pu leur donner naissance. Le cerveau et l'esprit ne font pas exception à la règle. Par conséquent, il nous faut comprendre la façon dont les structures cérébrales sous-jacentes à l'esprit sont apparues au cours de l'évolution des espèces.

Mais surtout, nous devons savoir comment fonctionnent ces structures. C'est ici que les progrès passionnants accomplis par les neurosciences prennent toute leur importance : il est extrêmement intéressant en effet de relier ces progrès aux réalisations des psychologues qui étudient le comportement et les processus mentaux.

Les résultats obtenus par les neurosciences permettent de penser que les processus mentaux résultent du fonctionnement de systèmes cérébraux extrêmement complexes à de nombreux niveaux d'organisation. Combien existe-t-il de niveaux ? Nous ne le savons pas vraiment, mais je mentionnerai certainement des niveaux moléculaires, des niveaux cellulaires, des niveaux organiques (concernant l'individu dans son ensemble) et des niveaux trans-organiques (c'est-à-dire des communications). Chaque niveau se subdivise encore, mais pour l'instant, je ne tiendrai compte que de ces divisions de base.

On est surpris de constater à quel point les connexions entre deux quelconques de ces niveaux sont nombreuses : entre la réaction de peur provoquée par un cri d'alarme et le processus chimique qui affectera les comportements ultérieurs ; entre une infection virale et une modification du développement cérébral qui influera sur la maturation ; entre la perception d'une forme et la chimie des modifications musculaires ; entre n'importe lequel de ces niveaux, à un moment critique du développement, et l'image de lui-même qu'un enfant se construira par la suite, qui peut être forte ou inadaptée, détachée ou dépendante selon les cas.

Pour pouvoir expliquer ces différents types de modifications, je dois tout d'abord dissiper un certain nombre d'idées erronées. Celles-ci sont dues principalement au fait que les représentants de diverses sous-disciplines sont restés confinés dans leur spécialité. Mais là n'est pas la seule raison. Les préjugés, l'incapacité de réaliser certaines expériences, et les pièges du langage expliquent la difficulté que l'on éprouve à démêler patiemment les liens entre phénomènes mentaux et phénomènes survenant dans le système nerveux.

Cependant, l'étude du problème de l'esprit est encore plus complexe que ne le laissent entrevoir ces éclaircissements. Nous verrons que les méthodes scientifiques qui s'appliquent aux objets inanimés, bien que fondamentales, ne sont pas adaptées à l'étude des animaux qui possèdent un cerveau et une intentionnalité. Cela est dû au fait que les observateurs scientifiques eux-mêmes sont des animaux ayant des intentions, enfermés dans leur propre expérience consciente. Ils doivent néanmoins s'assurer que leurs observations pourront être communiquées à d'autres observateurs d'une façon efficace, éloquente et exempte de préjugés. Cela signifie qu'ils ne peuvent pas – et même qu'ils ne doivent absolument pas – faire appel à des éléments de leur vécu personnel ou de leur conscience intime. Nous pouvons donc dire, en cherchant la rime (ou presque) :

en science, la communication intersubjective doit être objective et non projective. La magie, le vitalisme et l'animisme ont dominé la communication pré-scientifique. Non seulement la projection des souhaits, des croyances et des désirs personnels était autorisée, mais elle constituait même l'un des principaux objectifs des sociétés qui voulaient s'organiser pour défendre d'une manière rationnelle les individus contre les dangers naturels.

Pour autant, il n'est pas impossible d'étudier l'esprit d'un point de vue scientifique. Simplement, cette investigation sera nécessairement semée d'écueils, de présupposés, d'idées reçues – dont beaucoup seront issus de la science elle-même. Parmi les chercheurs qui étudient les propriétés de l'esprit, il arrive que même les plus intelligents trébuchent. Et lorsqu'ils étudient l'intentionnalité, certains d'entre eux persistent à pratiquer ce qui n'est qu'une parodie des sciences exactes telle la physique, c'est-à-dire des sciences consacrées à l'étude d'objets dépourvus d'intentionnalité.

Comment éviter de tomber dans ces pièges ? On peut tout d'abord les analyser et se demander si la recherche actuelle en neurosciences est à même de nous aider à les démanteler. Attelons-nous donc à cette tâche, efforçons-nous de replacer l'esprit dans la nature sans renoncer à la méthode scientifique.

Chapitre 2

Replacer l'esprit dans la nature

> *Évoquer la persécution de Galilée, c'est rendre hommage aux premiers commencements du changement d'attitude le plus radical que le genre humain ait jamais connu.*
>
> Alfred North Whitehead

Remontons dans le temps jusqu'à l'aube de la science moderne et considérons deux imposantes figures du XVII[e] siècle : Galileo Galilei et René Descartes. Dans *Science and The Modern World*, Alfred North Whitehead constate qu'en inventant la physique mathématique, Galilée a placé l'esprit hors de la nature. Je suppose qu'il voulait dire par là que Galilée insistait sur le fait que l'observateur doit être objectif, qu'il ne doit pas tomber dans les vaines disputes des philosophes aristotéliciens sur la question des causes. Au contraire, il se doit d'effectuer des mesures selon un modèle dépourvu de toute projection ou intention humaines, puis de rechercher les corrélations entre les phénomènes, ou lois, qui confirment ou infirment ces affirmations.

Cette façon de procéder a merveilleusement fonctionné aussi bien en physique que dans les sciences apparentées. Isaac Newton représente la figure triomphale de son premier grand épanouissement. Aujourd'hui encore, après la révolution einsteinienne et l'avènement de la mécanique quantique, la méthodologie galiléenne reste en vigueur. La théorie de la relativité d'Albert Einstein a montré comment la position et la vitesse de l'observateur modifient les mesures de l'espace et du temps et, en tenant compte de l'accélération, a opéré une modification du sens même du mot matière. Quant à la mécanique quantique, elle a permis de montrer que, dans le domaine de l'extrêmement petit, l'acte de mesure implique inéluctablement une intervention de l'observateur : dans les limites de l'incertitude imposée par la constante de Planck, celui-ci doit choisir le niveau de précision avec lequel il souhaite connaître la position ou la quantité de mouvement d'une particule subatomique.

Cela traduit ce que les physiciens appellent le principe d'incertitude d'Heisenberg.

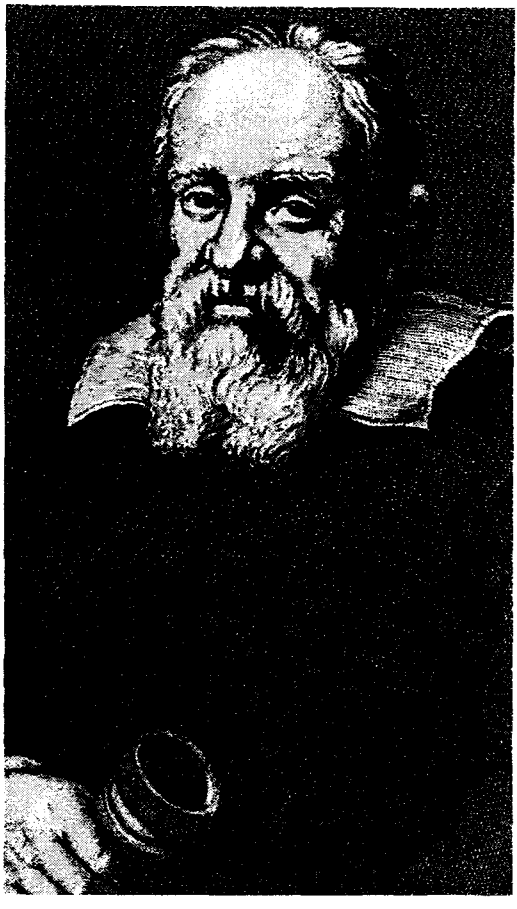


FIGURE 2-1

Galileo Galilei (1564-1642), physicien, mathématicien et astronome, fondateur de la physique moderne et, pour certains, de la science moderne. En 1633, l'Église catholique romaine l'obligea à renier la conception héliocentrique qu'il avait adoptée dans son Dialogue sur les deux grands systèmes du monde, *après quoi il passa le restant de sa vie assigné à résidence.*

En dépit des étonnantes révélations selon lesquelles à des vitesses proches de celle de la lumière ou à des distances extrêmement petites, l'observateur fait partie de ses propres mesures, l'objectif de la physique demeure galiléen : décrire des lois invariantes. Et de fait, il n'est pas de raison d'abandonner cet objectif, puisque les observateurs einsteiniens et heisenbergiens, même s'ils font partie de leurs propres mesures, demeurent psychologiquement transpa-

rents. Malgré les controverses soulevées parfois par les philosophes de la physique concernant leur importance dans les mesures quantiques, la conscience et les motivations de ces observateurs *n'ont pas besoin* d'être prises en considération quand on fait de la physique. L'esprit demeure bien séparé de la nature.

Mais, comme le faisait dûment remarquer Whitehead, l'esprit s'est trouvé replacé dans la nature au cours de la deuxième moitié du XIXe siècle, lors de l'avènement de la physiologie et de la psychologie physiologique. Et depuis, il n'a cessé de nous encombrer. De même que la relativité et la mécanique quantique ont quelque chose de singulier, les progrès de la physiologie posent des problèmes tout à fait particuliers : les observateurs eux-mêmes sont-ils des « choses », tout comme les autres objets qui peuplent leur monde ? Comment rendre compte de leur étonnante capacité (et même, pourrait-on dire, de leur besoin compulsif) de découper leur monde en différentes catégories d'objets – de se référer à des choses du monde alors que les choses elles-mêmes ne se réfèrent jamais à rien ? Lorsque à notre tour nous observons des observateurs, ce caractère intentionnel se révèle incontournable.

Devons-nous, pour rester dans la droite ligne de la physique, déclarer l'embargo sur tous les traits psychologiques dont nous parlons dans notre vie quotidienne – la conscience, la pensée, les croyances, les désirs ? Devons-nous adopter l'attitude minimaliste des behavioristes ? Les partenaires amoureux doivent-ils se dire : « C'était bon pour toi ; est-ce que c'était également bon pour moi ? » L'absurdité de ce dernier recours devient patente lorsqu'on considère la négation qu'il implique. Ou bien nous nions l'existence de ce que nous éprouvons (par exemple, de notre propre conscience) pour « devenir des chercheurs », ou bien nous déclarons que la science (la « science physique ») est incapable d'aborder ce genre de questions.

C'est ici que la deuxième grande figure de la révolution scientifique du XVIIe siècle, Descartes, réapparaît au premier plan. Dans sa quête d'une méthode de pensée, il a été conduit à professer un « dualisme des substances ». Selon ce point de vue, comme je l'ai déjà évoqué, le monde se compose de *res extensa* (substance étendue) et de *res cogitans* (substance pensante). Les manipulations galiléennes opèrent sur la *res extensa*, l'ensemble des choses étendues. Mais la *res cogitans*, l'ensemble des choses pensantes, n'existe pas à proprement parler dans le temps et l'espace ; dépourvue de lieu, n'étant pas une chose étendue, elle ne peut être accessible à l'ob-

servateur externe. Pire encore est le problème de l'interaction entre l'esprit et le corps, car ceux-ci doivent communiquer. Avec un manque de clarté qui ne lui était pas coutumier, Descartes déclara que les interactions entre la *res cogitans* et la *res extensa* avaient lieu dans la glande pinéale (figure 2-2).

La notion de dualisme a persisté, sous différentes formes, jusqu'à nos jours. Par exemple, tout en étant en apparence moniste, le behaviorisme n'est rien d'autre qu'un dualisme affaibli par le refus de considérer l'esprit comme un objet de science : c'est un dualisme auquel il manque un morceau. Les behavioristes résolvent le dilemme en étudiant les comportements sans tenir compte de l'intentionnalité. Ils ne tentent pas de replacer l'esprit dans la nature ; ils nient tout simplement sa validité en tant qu'objet d'études scientifiques. Par ailleurs, un grand nombre de psychologues non behavioristes, bien qu'ils affirment être matérialistes et non dualistes, n'en présupposent pas moins une dualité entre les propriétés de l'esprit et celles de la matière. En effet, tout en admettant que l'esprit et le cerveau sont issus d'une même substance, ils insistent sur le fait que les propriétés psychologiques doivent être traitées exclusivement dans des termes qui leur sont propres, et qui diffèrent

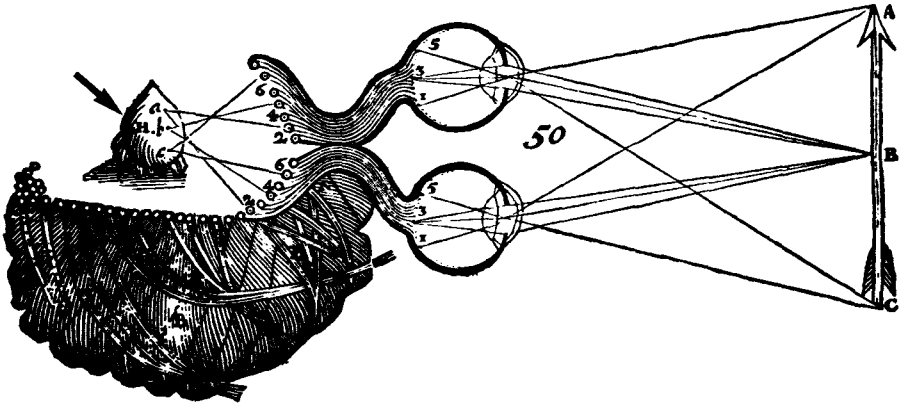


FIGURE 2-2

Diagramme du système visuel extrait du Traité de l'homme *de Descartes. Descartes suggéra que l'image rétinienne issue de chaque œil se projetait, par l'intermédiaire de fibres nerveuses, sur les parois des ventricules cérébraux remplis de liquide. La vision binoculaire était censée se projeter ensuite sur la glande pinéale* (flèche noire). *Descartes suggéra que cette structure sans pareille était le site sur lequel* la res cogitans *interagissait avec* la res extensa – *l'âme avec le corps. Nous savons à présent qu'en réalité la projection binoculaire s'effectue vers le cortex visuel, situé à l'arrière du cerveau (à la fois du côté droit et du côté gauche).*

nécessairement de l'angle sous lequel on considère les objets ou les corps physiques qui donnent lieu à ces propriétés. Vers la fin de sa vie, Sigmund Freud a bien incarné ce genre de dualisme.

Je voudrais faire remarquer que certains biologistes, même parmi les plus accomplis, ont éprouvé des doutes concernant l'idée d'étudier l'esprit. J'ai eu une fois l'occasion de discuter de ces questions avec l'éminent immunologiste Sir Peter Medawar au cours d'une conférence publique. Il se montra à peine sceptique : « À quoi ça sert ? », demanda-t-il. Je réussis à parer le coup en lui faisant remarquer que si nous comprenions mieux le cerveau, nous pourrions au moins nous débarrasser de quelques idées saugrenues sur la façon dont il fonctionne. Et, comme Peter était l'ennemi des explications jargonnantes, ma réponse le calma.

J'aurais aimé lui dire ce que Michael Faraday est censé avoir dit au ministre britannique des Finances William Gladstone, lorsqu'il lui présenta ses découvertes sur l'électricité. Comme ce monsieur lui demandait avec condescendance « À quoi ça sert ? », Faraday lui aurait répondu : « Un jour vous nous ferez payer des impôts dessus. » (En une autre occasion, il est censé avoir dit : « À quoi sert un nouveau-né ? »)

Dans notre tentative pour replacer l'esprit dans la nature, pouvons-nous dépasser le dualisme des substances et le dualisme des propriétés ? Ou allons-nous commettre de nouvelles erreurs ? Je suis obligé de nuancer ma réponse à ces deux questions. En effet, nous pouvons faire mieux, mais certainement pas en supposant, comme l'ont fait un certain nombre de spécialistes modernes de la cognition, que la structure et la biologie sont accessoires, qu'elles ne sont pas au centre de cette entreprise. Nous allons explorer cette question un peu plus en détail, car elle est riche de conséquences.

Au cours des dernières décennies, les chercheurs en sciences cognitives ont tenté, de manière sérieuse et approfondie, de transcender les limitations du behaviorisme. Les sciences cognitives relèvent d'un effort pluridisciplinaire en psychologie, en informatique et en intelligence artificielle. Elles font également appel à certaines données neurobiologiques, linguistiques et philosophiques. Certains chercheurs de chacune de ces disciplines, rendus plus audacieux par une apparente convergence d'intérêts, ont en effet choisi de ne pas rejeter d'emblée les fonctions mentales, comme le faisaient les behavioristes. Au contraire, ils se sont fondés sur la notion de représentation mentale et sur un ensemble d'hypothèses qualifiées de fonctionnalistes. Dans cette perspective, le comportement des

gens dépendrait de connaissances composées de représentations mentales symboliques, et la cognition consisterait à manipuler ces symboles. Les phénomènes psychologiques sont alors décrits en termes de processus fonctionnels. L'efficacité de ce type de processus réside dans la possibilité d'interpréter les choses comme des symboles abstraits et bien définis d'après un ensemble de règles sans équivoque. Un tel ensemble de règles constitue ce que l'on appelle une syntaxe.

L'application de ces règles de syntaxe est une forme de calcul [1]. Les calculs sont considérés comme largement indépendants de la structure et du mode de développement du système nerveux, tout comme un morceau de programme d'ordinateur peut être exécuté sur des machines dont les architectures sont différentes et est, par conséquent, « indépendant » de ces machines. Une idée voisine consiste à dire que le cerveau (ou plutôt l'esprit) est comme un ordinateur, que le monde est comme un morceau de bande magnétique, et qu'en général, le monde est si bien ordonné que les signaux reçus peuvent être « lus » en suivant un raisonnement logique.

De tels processus fonctionnels bien définis constitueraient des représentations sémantiques, ce qui veut dire qu'ils spécifieraient de façon non équivoque ce que leurs symboles représentent dans le monde. Sous sa forme la plus tranchée, cette vision des choses suggère que le substrat de toute activité mentale est en fait un langage de la pensée – un langage parfois appelé « mentalais » [2].

Ce point de vue – que certains appellent cognitiviste – est très en vogue et il a suscité, en psychologie, une profusion de travaux d'une grande valeur, qui ont mis en évidence nombre d'idées remarquables. Les êtres humains posséderaient, à la naissance, un dispositif d'acquisition du langage contenant les règles de syntaxe et de construction d'une grammaire universelle. La science (idéalement, la physique) serait capable de fournir une description non équivoque, objective, de la réalité. Cette description contribuerait à justifier les relations entre les processus ou règles syntaxiques et les choses ou événements – c'est-à-dire les relations constituant les représentations sémantiques. Enfin, le cerveau rangerait les objets du monde « réel » selon des catégories classiques, définies par des ensembles de conditions nécessaires et suffisantes.

---

1. À ce stade, je préfère ne pas donner de définition précise du mot calcul. Pour l'instant, disons qu'il consiste à manipuler des symboles selon une procédure précise. J'aborde cette question en détail dans la postface.
2. Voir la postface critique.

Je ne répéterai jamais assez à quel point ces idées, ou leurs variantes, ont envahi la science moderne. Elles sont généralisées et endémiques. Mais je dois également ajouter que l'entreprise cognitiviste repose sur un ensemble d'hypothèses qui n'ont jamais été analysées de près. L'un des plus étonnants déficits du cognitivisme est qu'il ne se réfère que de manière marginale aux bases biologiques sous-jacentes aux mécanismes qu'il prétend expliquer. Il en résulte une déviation scientifique aussi importante que celle du behaviorisme, qu'il prétend remplacer. Les erreurs cruciales qui sont à l'origine de cette déviation sont aussi mal perçues par la plupart des cogniticiens que ne l'était la relativité avant Einstein ou l'héliocentrisme avant Copernic.

Qu'est-ce exactement que les spécialistes ne réussissent pas à voir, et pourquoi est-ce si fondamental ? C'est l'idée que l'esprit ne peut pas être décrit de façon « pure » – c'est-à-dire en l'absence d'une description biologique détaillée du cerveau. Ces chercheurs négligent tout un ensemble de données qui vont à l'encontre de l'idée selon laquelle le cerveau serait une sorte d'ordinateur. Ils négligent les résultats qui montrent que la façon dont les êtres humains et les animaux classent les objets et les événements par catégories n'a rien à voir avec la logique ou le calcul. Et ils confondent la puissance formelle de la physique, telle qu'elle a été créée par des observateurs humains, avec la présomption selon laquelle les idées issues de la physique serviraient à comprendre des systèmes biologiques qui ont connu une évolution historique.

J'affirme que l'ensemble de la structure sur laquelle s'appuie le programme cognitiviste est incohérent et démenti par les faits. Je ne prétends pas étayer cette affirmation radicale dans le corps de ce livre ; pour ce faire, il me faudrait considérer de nombreuses disciplines et émettre de nombreuses hypothèses qui ne sont pas communément admises avant d'arriver à ma thèse principale. C'est pourquoi j'ai rassemblé mes arguments contre les différentes formes de cognitivisme pur au sein de la postface critique qui se trouve à la fin de cet ouvrage. Les spécialistes pourront consulter ce texte à loisir, mais je conseille aux lecteurs non spécialisés mais intéressés de ne s'y reporter qu'après avoir achevé la lecture du texte principal.

Cette postface aborde ce que je crois être une série d'erreurs de catégories. La première consiste à suggérer que la résolution des problèmes posés par la conscience viendra de la solution de certains dilemmes physiques. La deuxième consiste à envisager que l'informatique et l'intelligence artificielle puissent fournir les réponses.

La troisième, qui est aussi la plus énorme, est l'idée que pour faire aboutir l'ensemble de l'entreprise, il suffit d'étudier les comportements, les performances et les aptitudes mentales, ainsi que le langage, en partant simplement des hypothèses fonctionnalistes, donc sans comprendre d'abord la biologie sous-jacente.

J'aborderai les arguments essentiels dans la postface. Au cours du chapitre qui suit, je me propose de survoler un certain nombre de faits et d'idées issus de la biologie et des neurosciences. Il est vital de comprendre quelle est cette matière qui sous-tend effectivement l'esprit, et en particulier les principes qui président à son organisation. Seule cette connaissance nous permettra de disséquer les difficultés auxquelles nous sommes confrontés lorsque nous tentons d'étudier l'esprit, et de proposer un certain nombre d'issues aux situations fâcheuses que j'ai évoquées ci-dessus.

Je partirai du principe suivant : il doit exister des manières de replacer l'esprit dans la nature qui correspondent justement à la façon dont l'esprit est initialement apparu dans la nature. Elles doivent tenir compte de ce que nous apprend la théorie de l'évolution. En effet, c'est au cours de l'évolution que les corps en sont venus à posséder des esprits. Mais il ne suffit pas de dire que l'esprit a pris corps, qu'il s'est incarné ; il faut réussir à préciser comment. Et pour ce faire, nous devons examiner non seulement le cerveau et le système nerveux, mais aussi les problèmes structurels et fonctionnels qu'ils présentent.

Chapitre 3

La matière de l'esprit

Les seules lois de la matière sont celles que nos esprits doivent fabriquer, et les seules lois de l'esprit sont fabriquées pour lui par la matière.

James Clerk Maxwell

Étant donné le caractère unique de la conscience et le fait que la pensée est incapable de « scruter » ses propres mécanismes, il n'est pas étonnant que certains philosophes aient suggéré qu'il existe une substance pensante. Certains ont même imaginé une sorte de panpsychisme selon lequel toute matière posséderait une conscience. Cependant, les résultats de la science moderne incitent à penser que la matière physique sous-jacente à l'esprit n'a rien de spécial. Elle est tout à fait ordinaire – autrement dit, elle est faite d'éléments chimiques tels que le carbone, l'hydrogène, l'oxygène, l'azote, le soufre et le phosphore, ainsi que d'un certain nombre de métaux à l'état de trace. Par conséquent, il n'y a rien dans la composition de base du cerveau qui puisse nous renseigner sur la nature des propriétés mentales.

Ce qui est particulier au cerveau, c'est son organisation. Ces éléments chimiques ordinaires forment un ensemble de molécules extraordinairement complexes, qui à leur tour constituent les structures cellulaires complexes des tissus vivants. Dans un organisme aussi élaboré qu'un être vivant, il existe environ deux cents types différents de cellules. Parmi celles-ci, les cellules nerveuses, ou neurones, constituent l'un des types cellulaires les plus spécialisés et les plus exotiques. L'originalité des neurones est triple : elle réside dans la variabilité de leur forme, dans leurs fonctions électriques et chimiques, et dans leur connectivité – c'est-à-dire dans la façon dont les neurones s'interconnectent pour former des réseaux.

J'ai l'intention de vous en dire plus sur certaines de ces propriétés, juste assez pour vous convaincre du fait que nous nous trouvons devant quelque chose qui ne ressemble à rien d'autre dans l'univers. Je compléterai cette description à mesure que cela deviendra nécessaire, ce qui nous permettra de ne pas nous encombrer dès le début

de données complexes. Mais avant de vous exposer un certain nombre de détails descriptifs, il me semble utile de vous donner une idée du nombre de neurones que l'on trouve dans certaines régions du cerveau et du nombre de connexions qu'ils établissent entre eux. Cela constituera, à mon avis, un point de départ suffisamment surprenant. Cependant, quand j'en viendrai à présenter un peu de morphologie, vous serez peut-être encore plus impressionné par ce que l'évolution a réussi à accomplir en sélectionnant les animaux pour la richesse de leur structure cérébrale.

Commençons par la partie du cerveau appelée cortex cérébral (figure 3-1). Cette structure se trouve au centre de ce que l'on appelle d'un terme vague les fonctions cérébrales supérieures – la parole, la pensée, les mouvements complexes, la musique. Si on étalait la totalité de ce « manteau » ondulé, qui recouvre le haut et les côtés de notre cerveau, il aurait la taille d'une grande serviette de table et à peu près la même épaisseur. Bien que l'on ne connaisse pas avec précision le nombre de cellules présentes dans cette structure, on estime qu'elle possède dix milliards de neurones. (On y trouve également des cellules de support, les cellules gliales, mais je n'en tiendrai pas compte.)

Chaque cellule nerveuse établit des connexions avec d'autres cellules nerveuses au niveau de sites appelés synapses. Et, fait étonnant, il y a environ un million de milliards de connexions dans la couche corticale. Au rythme d'une connexion (ou synapse) par seconde, il vous faudrait quelque trente-deux millions d'années pour les compter toutes. Pour sentir à quel point le nombre de connexions au sein de cette extraordinaire structure est grand, on peut aussi considérer qu'un morceau de notre cerveau de la taille d'une grosse tête d'allumette contient environ un milliard de connexions. Vous remarquerez que je n'ai évoqué ici que le nombre de connexions dans un cerveau donné. En effet, si nous considérions toutes les connexions et toutes leurs combinaisons possibles, nous obtiendrions un nombre astronomique – de l'ordre d'un dix suivi de millions de zéros. (Il n'y a, dans tout l'univers connu, qu'environ 10 à la puissance 80 particules chargées positivement !)

Voilà donc un premier indice suggérant que le cerveau est si singulier qu'il a bel et bien pu donner lieu aux propriétés mentales. Cependant, bien que le nombre et la densité même des réseaux neuronaux du cerveau soient spectaculaires, ce ne sont pas là les seules propriétés du tissu cérébral. Une propriété encore plus remarquable tient à la façon dont les cellules cérébrales sont organisées

en structures fonctionnelles. Quand on considère à la fois ce superbe agencement cellulaire (la micro-anatomie, ou morphologie, cérébrale), le nombre de cellules contenues dans le cerveau et les réactions chimiques qui s'y déroulent, il est clair qu'il s'agit là de l'objet matériel le plus complexe que nous connaissions dans l'univers.

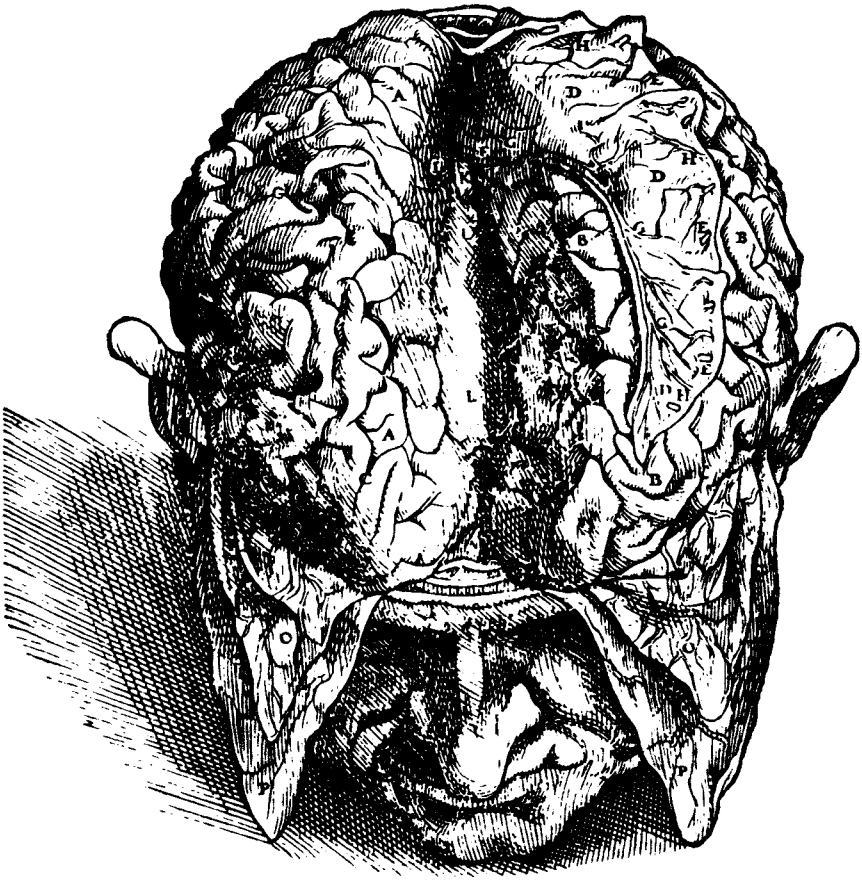


FIGURE 3-1

La surface du cortex cérébral humain, dessinée par le grand anatomiste André Vésale (1514-1564). Vésale est généralement considéré comme le père de l'anatomie moderne, et son De Fabrica Humanis Corpora *a servi à définir de nouvelles normes pour l'art médical.*

Je voudrais en dire un peu plus sur quelques-unes des propriétés des autres composants du cerveau. Chez des animaux aussi complexes que les êtres humains, le cerveau comporte des couches et des structures plus ou moins arrondies, appelées noyaux. Chacune de

ces structures a évolué de manière à être capable d'assurer des fonctions au sein d'un réseau complexe de connexions, et chacune comporte un très grand nombre de neurones – parfois plus et parfois moins que le cortex. Le cerveau est relié au monde extérieur par l'intermédiaire de neurones spécialisés appelés transducteurs sensoriels, qui composent les organes sensoriels et fournissent des entrées au cerveau. Quant aux sorties produites par celui-ci, elles sont reliées aux muscles et aux glandes par d'autres neurones. Mais certaines parties du cerveau (en fait, la majeure partie de ses tissus) ne reçoivent que des entrées provenant d'autres parties du cerveau et envoient leurs sorties à d'autres parties sans aucune intervention du monde extérieur. Donc, pour ainsi dire, le cerveau est davantage en contact avec lui-même qu'avec n'importe quoi d'autre.

Comment les neurones se lient-ils les uns aux autres et comment sont-ils organisés au sein des noyaux et des couches ? Comme je l'ai déjà évoqué, les synapses constituent le principal moyen de connexion. Une synapse est une structure spécialisée dans laquelle l'activité électrique transmise le long de l'axone du neurone pré-synaptique (figure 3-2) induit la libération d'une substance chimique (appelée neuromédiateur) qui, à son tour, induit une activité électrique dans le neurone post-synaptique. Comme on le voit sur la figure, la force ou efficacité des synapses peut être modifiée : au niveau pré-synaptique par des modifications de la quantité de neuromédiateur libérée, et au niveau post-synaptique par la modification de l'état chimique des récepteurs et des canaux ioniques – les unités qui respectivement, du côté post-synaptique, se lient aux médiateurs et laissent pénétrer les ions, porteurs de charges électriques (comme par exemple les ions calcium), à l'intérieur de la cellule.

Les neurones peuvent prendre des formes variées, et c'est cette forme qui détermine en partie la façon dont un neurone se lie aux autres, engendrant ainsi la neuro-anatomie d'une région cérébrale donnée. Du point de vue anatomique, les neurones peuvent être organisés de plusieurs façons différentes. Parfois, ils forment des cartes. L'établissement de cartes constitue un principe d'organisation important dans les cerveaux complexes. Les cartes font correspondre des points situés sur les diverses couches réceptrices bidimensionnelles du corps (telles la peau ou la rétine de l'œil) à des points situés sur les couches qui composent le cerveau. Les couches réceptrices (par exemple, les cellules tactiles qui se trouvent au bout de nos doigts

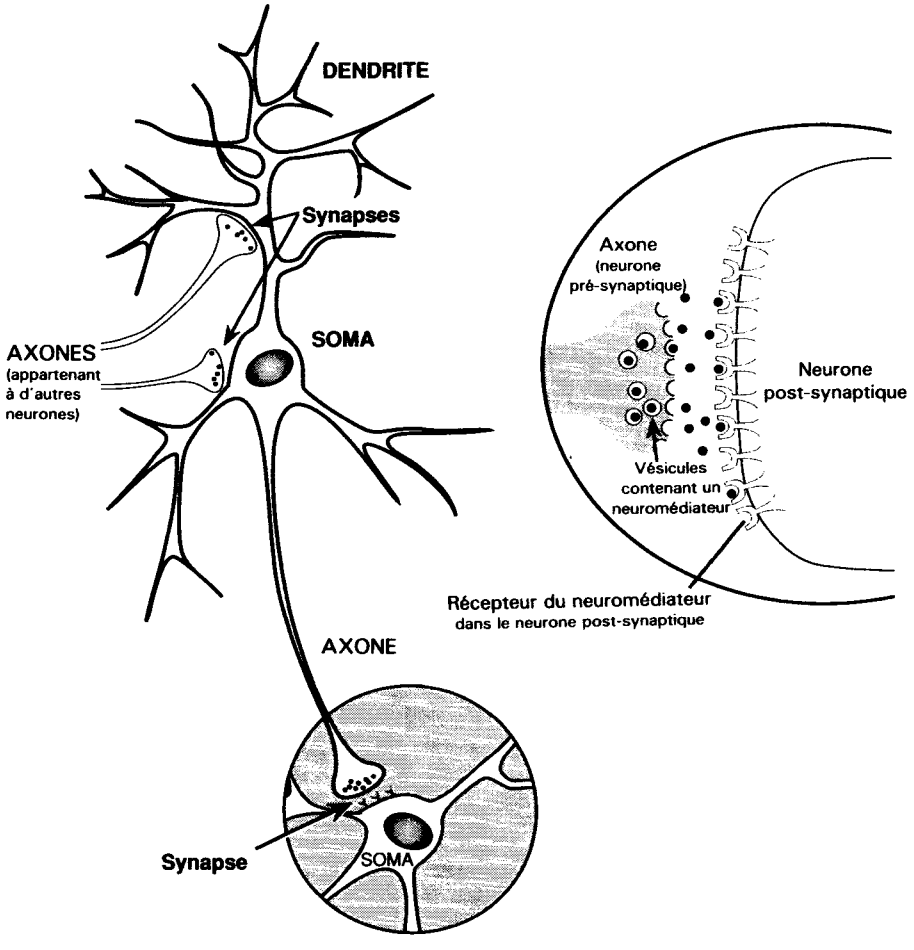


FIGURE 3-2

Plusieurs organisations de la matière de l'esprit, illustrées par des neurones simplifiés. Les axones provenant de neurones proches ou lointains sont de longs prolongements qui établissent des contacts avec d'autres neurones, soit au niveau du corps, ou soma, de ceux-ci, soit au niveau de leurs ramifications, appelées dendrites. Les axones transportent l'activité électrique qui entraîne la libération d'un neuromédiateur lorsqu'elle atteint la synapse entre deux neurones. Après avoir interagi avec les récepteurs correspondants, le neuromédiateur stimule à son tour le neurone récepteur (ou post-synaptique), qui se met à décharger de l'électricité. On voit une synapse simplifiée dans le cercle situé dans la partie inférieure de cette figure, et un agrandissement de cette synapse dans le cercle situé à droite. Les petits sacs ronds représentent des vésicules dans lesquelles le neuromédiateur est stocké dans le neurone pré-synaptique. Le neurone post-synaptique arbore ses récepteurs (les structures en forme de Y) dans l'espace qui sépare les membranes pré-synaptique et post-synaptique. Ces récepteurs se lient au médiateur libéré par les vésicules pré-synaptiques et déclenchent les réponses du neurone post-synaptique.

et les cellules rétiniennes sensibles à la lumière) sont capables de réagir au monde tridimensionnel et de fournir au cerveau des informations spatiales concernant les différences de pression ou de longueur d'onde (les récepteurs réagissent en fait à un monde de dimension quatre si l'on tient également compte du temps). En outre, les cartes cérébrales sont reliées entre elles via des fibres qui sont plus nombreuses que tout autre type de fibres dans le cerveau. Ainsi par exemple, le corps calleux – le principal faisceau de fibres qui, à travers la ligne médiane du cerveau, relie des parties de l'hémisphère droit à des parties de l'hémisphère gauche – contient environ deux cents millions de fibres.

Rien de tout cela n'était connu avant le XIX[e] siècle. Mais des hypothèses avaient déjà été formulées avant cette époque par des hommes remarquables, Denis Diderot par exemple. Voici un extrait de son roman en forme de pièce de théâtre, *Le Rêve de d'Alembert*, dans lequel la maîtresse de d'Alembert, Mademoiselle de l'Espinasse, s'enquiert auprès d'un médecin, le docteur Bordeu, des causes des étranges rêves de d'Alembert.

BORDEU : C'est qu'il n'en est pas du tic de l'origine comme du tic d'un des brins. La tete peut bien commander aux piés, mais non pas le pié a la tete ; l'origine a un des brins, non pas le brin a l'origine.

MADEMOISELLE DE L'ESPINASSE : Et la difference, s'il vous plait ? En effet, pourquoi ne pensé-je pas partout ? C'est une question qui auroit du me venir plutot.

BORDEU : C'est que la conscience n'est qu'en un endroit.

MADEMOISELLE DE L'ESPINASSE : Voila qui est bientot dit.

BORDEU : C'est qu'elle ne peut être que dans un endroit, au centre commun de toutes les sensations, la ou est la memoire, la ou se font les comparaisons. Chaque brin n'est susceptible que d'un certain nombre determiné d'impressions, de sensations successives, isolées, sans memoire. L'origine est susceptible de toutes, elle en est le registre, elle en garde la memoire ou une sensation continue, et l'animal est entrainé dès sa formation premiere a s'y raporter soi, a s'y fixer tout entier, a y exister.

MADEMOISELLE DE L'ESPINASSE : Et si mon doigt pouvoit avoir de la memoire ?...

BORDEU : Votre doigt penseroit.

MADEMOISELLE DE L'ESPINASSE : Et qu'est-ce donc que la memoire ?

BORDEU : La proprieté du centre, le sens specifique de l'origine du reseau, comme la vue est la propriété de l'œil ; et il n'est pas plus etonant que la memoire ne soit pas dans l'œil, qu'il ne l'est que la vue ne soit pas dans l'oreille.

MADEMOISELLE DE L'ESPINASSE : Docteur, vous eludez plutot mes questions que vous n'y satisfaites.

BORDEU : Je n'elude rien, je vous dis ce que je scais, et j'en scaurois davantage, si l'organisation de l'origine du reseau m'etoit aussi connue que celle de ses brins, si j'avois eu la meme facilité de l'observer. Mais si je suis foible sur les phenomenes particuliers, en revanche, je triomphe sur les phenomenes généraux.

MADEMOISELLE DE L'ESPINASSE : Et ces phenomenes generaux sont ?

BORDEU : La raison, le jugement, l'imagination, la folie, l'imbecillité, la ferocité, l'instinct.

[...]

BORDEU : Et les habitudes qui subjuguent ; le vieillard qui aime les femmes, et Voltaire qui fait encore des tragedies.

(En cet endroit le docteur se mit à rever et mademoiselle de l'Espinasse lui dit :) Docteur, vous revez.

BORDEU : Il est vrai.

MADEMOISELLE DE L'ESPINASSE : A quoi revez-vous ?

BORDEU : A propos de Voltaire.

MADEMOISELLE DE L'ESPINASSE : Hé bien ?

BORDEU : Je reve a la manière dont se font les grands hommes.

À ce stade de nos réflexions, nous disposons déjà d'un certain nombre de faits sur la matière de l'esprit. Nous savons que les cellules nerveuses sont spécialisées, qu'elles sont nombreuses, et que leurs connexions, qui sont hyperdenses, possèdent à leur tour des caractéristiques chimiques et morphologiques particulières. L'anatomie qui résulte de cette organisation est spectaculaire par son imbrication et sa variété. Mais elle dépend également de certains principes d'organisation généraux : elle est faite de couches, contenant des cartes topographiques, et de noyaux arrondis (des *blobs* en anglais). Elle projette aussi de nombreuses fibres chargées de relier les cartes aux couches sensorielles, et d'autres vers les muscles du corps. Enfin, chaque carte contient des cartes d'autres cartes.

La stimulation des éléments sensoriels induit, dans la membrane des neurones, des signaux nerveux sous la forme de décharges électriques. Ces décharges sont engendrées par le flux d'ions chargés. (Cela signifie que les charges électriques se déplacent plus lentement dans les membranes cellulaires que dans les fils du téléphone, où le courant est porté par des électrons.) Agissant en parallèle, des batteries de neurones donnent ainsi lieu à un nombre étonnamment élevé de combinaisons différentes. En outre, leur sensibilité aux stimulus peut être modifiée par une multitude de substances

chimiques – notamment par les neuromédiateurs au niveau des synapses, par d'autres substances, appelées neuromodulateurs, et, bien entendu, par des médicaments.

Un fragment de tissu cérébral est un réseau imbriqué qui réagit aux signaux électriques et chimiques dans l'espace tridimensionnel et dans le temps. Il émet des signaux dynamiques, en reçoit d'autres et y réagit. Ces signaux influent les uns sur les autres et, à travers d'autres connexions nerveuses, ils influent également sur l'action des autres organes du corps – tels le cœur, les reins, les poumons, les muscles et les glandes. Le cerveau est le centre de commandes et ses signaux rythmiques modifient notre façon de respirer, de pomper du sang, de digérer la nourriture, de nous déplacer.

J'aborderai, dans un chapitre ultérieur, les principes en vertu desquels le système nerveux (et, en fait, l'animal tout entier) acquiert sa forme, aussi bien à l'échelle microscopique que globale. Mais il est utile d'anticiper un peu ici sur cette discussion. Les agencements anatomiques du cerveau et du système nerveux résultent d'une série d'événements qui surviennent au cours du développement de l'organisme (figure 3-3). Chez l'embryon, les cellules se divisent, migrent, meurent, adhèrent les unes aux autres, projettent des excroissances et établissent (ou abolissent) des connexions synaptiques. Cette séquence dynamique d'événements dépend très fortement de la position dans l'espace des cellules (c'est-à-dire des cellules qui se trouvent aux alentours), du temps (c'est-à-dire de l'instant auquel un événement survient par rapport à un autre) et de la corrélation de l'activité cellulaire (le fait qu'au cours d'un certain laps de temps, des cellules envoient ou non simultanément des décharges électriques, ou qu'elles subissent ou non simultanément des modifications chimiques).

Au cours du développement, les dépendances spatiales sont frappantes, les plus nettes apparaissant lors de la formation des cartes pendant le développement embryonnaire. Dans le cas, par exemple, de la carte de l'espace visuel, formée par les projections rétinotectales (figure 3-4), les neurites (fibres) issues des cellules ganglionnaires de la rétine forment le nerf optique. Elles se projetteront ensuite, d'une façon très précise du point de vue topographique, sur une région qui, chez un animal comme la grenouille, porte le nom de toit optique. La stimulation d'un point donné de la rétine par un point lumineux induit la stimulation des neurones d'une région donnée du toit optique, et les cellules stimulées forment une carte

bien définie. L'organisation spatiale des cellules est indispensable au bon fonctionnement d'une telle carte.

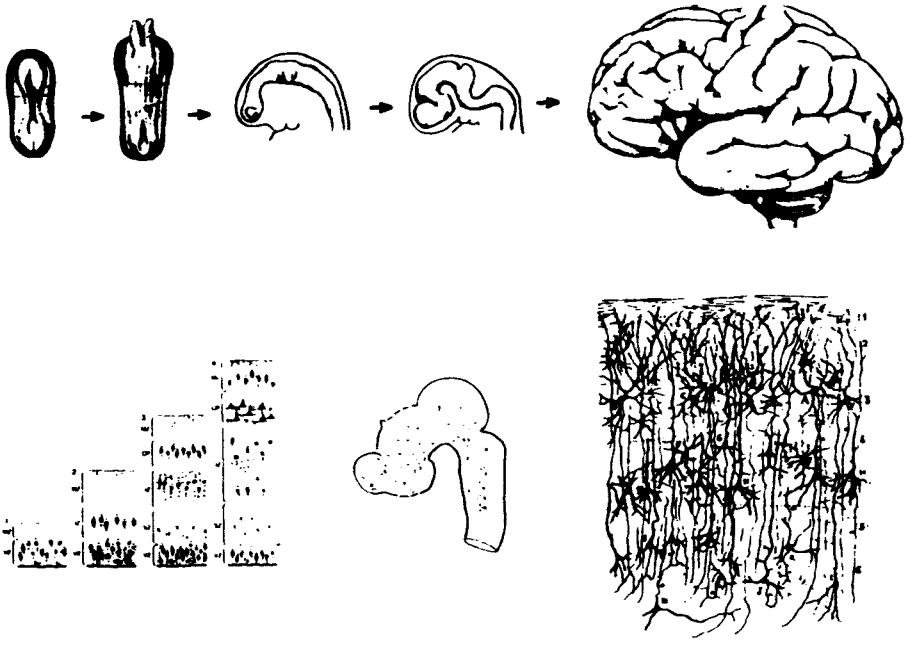


FIGURE 3-3

Le développement du cerveau – de la gouttière neurale (en haut à gauche) *au cortex cérébral* (en haut à droite). *Les neurones forment des couches* (en bas à gauche), *ou bien ils suivent divers parcours* (en bas au centre), *et finissent par interagir à travers les synapses pour donner lieu à une neuro-anatomie extrêmement complexe* (en bas à droite). *À un moment ou un autre de leur carrière, tous les neurones sont des nomades : ils se déplacent vers leurs positions finales sur d'autres cellules. Il en résulte l'objet matériel le plus complexe que l'on connaisse dans l'univers.*

Au cours du développement, l'organisation de cette carte est obtenue au bout d'au moins deux étapes. La première étape, qui fait intervenir la croissance de neurites à partir des fibres du nerf optique par arborisations qui se recouvrent partiellement, aboutit à la formation d'une carte grossière et ne requiert aucune activité neuronale. La seconde étape, au cours de laquelle la carte est affinée et devient beaucoup plus précise, requiert que l'activité dans les fibres de cellules ganglionnaires voisines soit corrélée avec l'activité dans le toit optique. La formation des cartes durant le développement chez des animaux comme le poisson rouge ou la grenouille est dynamique ; les connexions se déplacent et se rétablissent à mesure que la croissance différentielle se déroule à la fois dans la

rétine et le toit optique. Les principes qui régissent ces modifications sont épigénétiques – autrement dit, les événements clés ne se produisent que si certains autres événements ont eu lieu au préalable. L'une des conséquences importantes de ce phénomène est que les connexions entre cellules ne sont pas spécifiées d'avance de manière précise dans les gènes de l'animal.

Ce qui rend les cartes si intéressantes, c'est que les événements épigénétiques qui engendrent la forme à partir de la position dans l'espace au cours des premiers stades du développement embryonnaire doivent, dans une certaine mesure, « prévoir » les interactions futures des surfaces bidimensionnelles des couches de récepteurs sensoriels (la rétine ou la peau, par exemple) avec le monde tridimensionnel dans lequel l'animal devra se mouvoir et recevoir des stimulus. Nous verrons plus loin comment les principes du développement et de l'évolution rendent compte de ces phénomènes.

Jusqu'à présent, ma description peut ressembler à celle d'un vaste central téléphonique, ou peut-être même à celle d'un ordinateur numérique. Et, d'une certaine manière, le cerveau se comporte effectivement comme ce genre de systèmes. Cependant, lorsque nous examinons en détail les caractéristiques structurelles et les propriétés fonctionnelles du système nerveux, l'analogie s'effondre et nous sommes confrontés à une série de problèmes. Ces problèmes ont donné lieu à une série de crises d'interprétation, aussi bien en neurosciences que dans les sciences qui en sont tributaires.

Les crises structurelles, que j'ai déjà décrites de façon détaillée dans mon livre intitulé *Neural Darwinism*, concernent l'anatomie et le développement. En effet, bien qu'à une certaine échelle, le cerveau ressemble à un vaste réseau électrique, à son échelle la plus microscopique, ses connexions et son organisation ne ressemblent à celles d'aucun autre réseau, qu'il soit naturel ou construit par l'homme. Comme nous venons de le voir, le réseau cérébral est engendré au cours du développement par des déplacements cellulaires et par l'extension et l'interconnexion d'un nombre croissant de neurones. Autrement dit, le cerveau est un exemple de système capable de s'auto-organiser. Et l'examen de ce système au cours de son développement, ainsi que celui de ses ramifications les plus microscopiques à l'issue du développement, indiquent que la mise en place d'un câblage précis, point par point (comme celui d'un dispositif électronique), y est impossible. Les fluctuations observées sont bien trop grandes.

De plus, bien que la connectivité des systèmes neuronaux dans

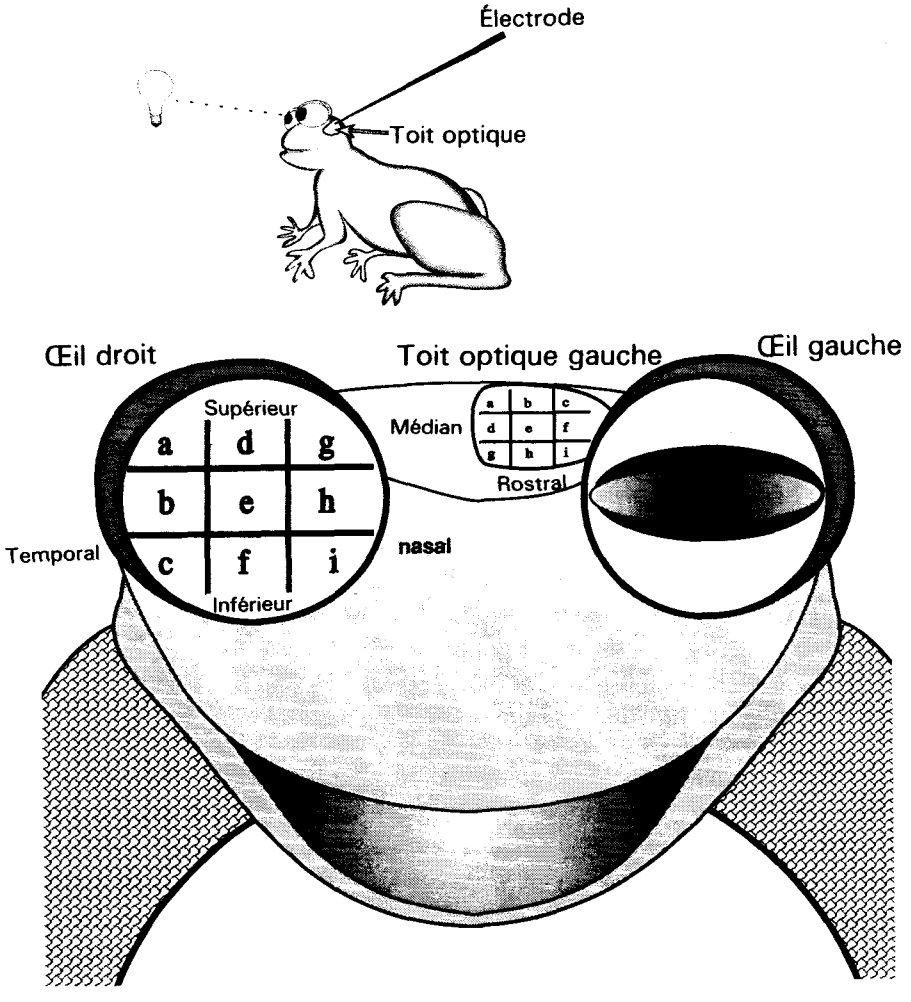


FIGURE 3-4

Cartographie de l'œil et de ses champs visuels sur le cerveau. En haut : *une grenouille avec une électrode dans la partie visuelle de son cerveau (appelée toit optique).* En bas : *vue de face. Les parties de la rétine de l'œil droit indiquées par des lettres ont chacune été éclairées pendant que l'on mesurait les réponses électriques dans le toit optique de la grenouille à l'aide d'une électrode. Dans le toit optique gauche, les régions voisines correspondantes forment une carte de réponses, indiquées par les mêmes lettres. Remarquez que, bien que la carte ait subi une rotation, les voisins demeurent les mêmes dans l'œil et du côté opposé du toit optique. Les régions temporales de l'œil se projettent sur les régions médianes du toit optique, les régions inférieures sur les régions rostrales, et ainsi de suite.*

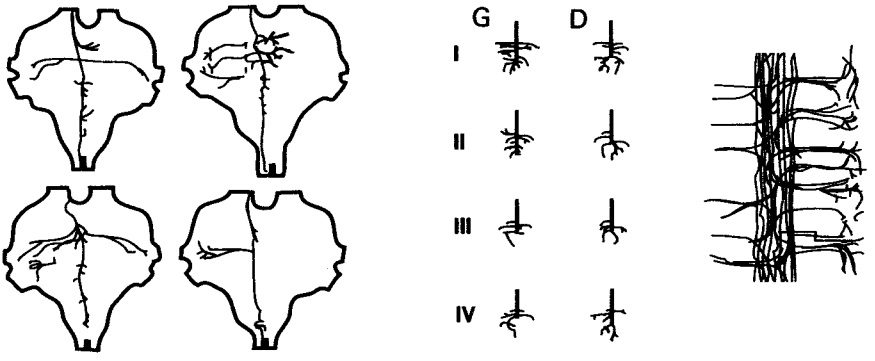
le système nerveux central (et en particulier de ceux qui donnent lieu à des cartes) soit plus ou moins semblable d'un individu à un autre, elle n'est pas identique. De fait, comme le montre la figure 3-5, la forme des neurones de chaque classe, tout comme l'organisation de leurs connexions, varient considérablement. Cela n'est pas surprenant, vu la nature stochastique (ou statistiquement variable) des forces qui régissent le développement, et qui sont fournies par des processus cellulaires tels que la division, les déplacements et la mort cellulaires. Dans certaines régions du système nerveux en cours de développement, il arrive que jusqu'à 70 % des neurones meurent avant que la structuration de la région ne soit achevée ! Par conséquent, en général, il ne peut exister de connexions spécifiées de façon figée. Si nous comptions les ramifications d'un neurone donné, puis les neurones atteints par ces ramifications, les deux nombres obtenus ne seraient pas exactement identiques d'un individu à un autre de la même espèce – et cela même s'il s'agissait de jumeaux identiques ou d'animaux génétiquement identiques.

Pour compliquer encore les choses, les neurones projettent en général les ramifications de leurs axones sous forme d'arborisations divergentes qui *recouvrent partiellement* celles qui proviennent d'autres neurones. Il en va de même pour les excroissances, appelées dendrites, des neurones receveurs (voir la figure 3-2). J'ai déjà illustré cette situation lorsque j'ai abordé la question des arborisations des fibres du nerf optique qui se projettent sur le toit optique. Pour dire les choses de manière plus imagée, si nous « demandions » à un neurone quel est celui qui, parmi tous ceux qui sont reliés à l'ensemble de ses dendrites, lui a fourni une entrée donnée, il n'aurait aucun moyen de le « savoir », parce que ses dendrites se recouvrent partiellement.

Le fait que certains principes du développement introduisent une variabilité dans les connexions, et qu'ils permettent à des arborisations de se chevaucher – ce qui donne lieu à une organisation synaptique non identifiable (et peut-être même non reproductible) –, a engendré une crise chez ceux qui croient que le système nerveux est précis et « câblé » à la manière d'un ordinateur. Nous sommes donc en droit de nous demander comment, lorsqu'elle a été perçue, cette crise a été résolue par ceux qui croient que le cerveau ressemble à un ordinateur.

En premier lieu, ceux qui adhèrent à cette idée négligent toute variabilité au-dessous d'un certain niveau microscopique en allé-

VARIABILITTÉ STRUCTURELLE



VARIABILITÉ DYNAMIQUE

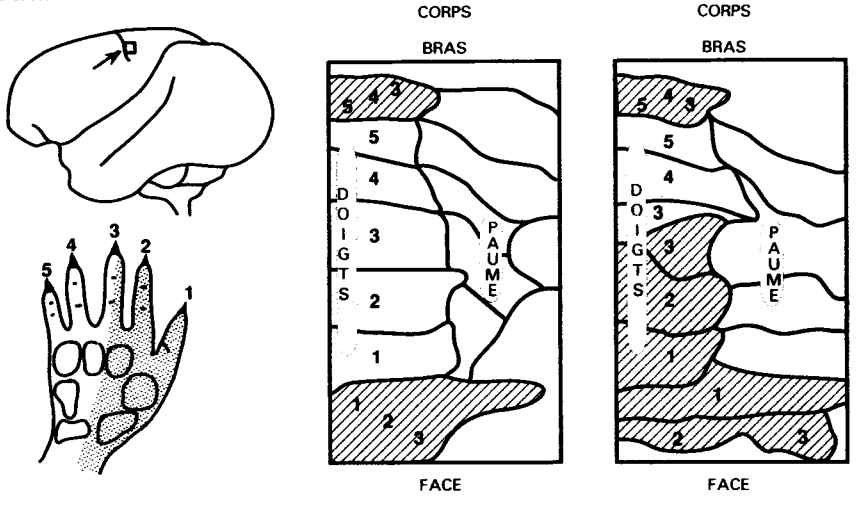


FIGURE 3-5

Variabilité des structures neuronales. En haut à gauche : *la structure du même nerf chez quatre criquets* différents *(travail réalisé par Keir Pearson et Corey Goodman).* En haut, au centre : *des neurones visuels provenant de quatre puces d'eau différentes mais génétiquement identiques (visualisés par Eduardo Macagno et ses collègues). Remarquez que même les neurones qui se correspondent du côté gauche et du côté droit d'un même individu ne sont pas identiques.* En haut à droite : *ces structures nerveuses répétitives, provenant du cerveau d'un lapin, sont toutes différentes (voir le côté droit).* En bas : *le dynamisme et la variabilité des cartes cérébrales correspondant au toucher chez un singe de nuit adulte. On voit au milieu une carte normale (emplacement indiqué par la flèche dans le cerveau figuré à gauche) des doigts, de la paume (régions claires numérotées), et des régions correspondantes du dos, ou côté poilu, de la main (régions sombres). Après section du nerf qui dessert une partie du côté avant (dépourvu de poils) des doigts et la paume, une réorganisation des frontières de la carte se produit – non seulement au niveau des zones correspondant à l'avant et à l'arrière de la main, mais aussi au niveau des zones correspondant aux autres doigts (carte à droite). (Ce travail a été réalisé par Michael Merzenich et ses collègues.)*

guant qu'il ne s'agit là que de « bruit » – l'une des conséquences inévitables du dilemme du développement. Deuxièmement, ils rendent compte de l'absence de connexions spécifiées de manière unique en arguant que les structures hautement organisées – les cartes, par exemple – ne requièrent pas de telles connexions ou compensent leur absence d'une manière ou d'une autre. Et troisièmement, ils expliquent l'absence d'entrées synaptiques précisément identifiées en supposant que les neurones utilisent un code semblable à ceux qu'on utilise pour identifier les cartes de crédit téléphoniques ou les utilisateurs d'un ordinateur. Dans les neurones, les codes de position et de temps sont censés dépendre de la fréquence, de l'espacement ou du type d'activité électrique, ou encore des types de médiateurs chimiques auxquels ces neurones sont associés (figure 3-2). Remarquons, cependant, que ces explications présupposent que chaque neurone pris isolément est porteur d'*information*, à l'instar de certains dispositifs électroniques. Je montrerai plus loin que cette hypothèse n'est pas tenable et que ces explications sont inadéquates. On n'a jamais trouvé, dans le système nerveux humain, de preuve convaincante de l'existence de codes du type de ceux que les êtres humains utilisent en télégraphie, en informatique ou dans d'autres formes de communication de ce type.

Cela nous amène à proposer un certain nombre d'énigmes plus profondes à ceux qui suggèrent que le cerveau est une sorte d'ordinateur. Ces énigmes constituent un ensemble de crises fonctionnelles relevant de la physiologie et de la psychologie. La première énigme est la suivante : lorsque nous explorons le réseau microscopique de synapses avec des électrodes afin de détecter les résultats des décharges électriques, nous constatons que la plupart des synapses ne s'expriment pas, c'est-à-dire ne présentent aucune décharge détectable. Il s'agit de ce qu'on appelle des « synapses silencieuses ». Mais pourquoi sont-elles silencieuses, et quel est le rapport entre leur silence et les signaux, codes ou messages qu'elles sont censées porter ?

Un deuxième dilemme concerne les fonctions et les interactions des cartes du type de celles que nous avons déjà évoquées dans le cas du système rétinotectal. En effet, malgré ce que nous dicte la sagesse conventionnelle des livres d'anatomie, ces cartes ne sont pas figées ; dans certaines aires du cerveau, on assiste à des fluctuations extrêmement marquées au niveau des frontières des cartes au cours du temps. De plus, les cartes de chaque individu semblent être uniques. Plus frappant encore, la variabilité des cartes chez

les animaux adultes dépend de celle des entrées disponibles pour la réception des signaux (figure 3-5). À première vue, cela pourrait paraître ne poser aucun problème ; après tout, les ordinateurs modifient également leurs « cartes », ou tableaux, lorsqu'on modifie le logiciel. Mais les cartes *opérationnelles* du système nerveux sont fondées sur des cartes *anatomiques* – et, à ce niveau anatomique, dans le cerveau adulte, il n'y a que la mort des neurones qui soit capable de les modifier. En outre, même si l'on suppose que ces cartes se modifient sous l'effet de modifications « logicielles », on a du mal à voir quel pourrait bien être le code qui ferait que deux individus dont les cartes anatomiques ne sont pas identiques aboutissent à la même sortie, au même résultat. On l'explique en général en disant qu'il existe dans le cerveau des systèmes alternatifs pour traiter les entrées changeantes : chacun de ces systèmes est fixe et câblé une fois pour toutes, mais pourrait être activé ou inactivé en fonction des entrées. Les faits montrent, cependant, que la variabilité des cartes neuronales n'est pas discrète ou binaire, mais plutôt qu'elle est constante, qu'elle est présente à des échelles très fines, et qu'elle est très étendue. Par conséquent, il faudrait que le nombre de cartes alternatives soit très grand.

D'autres observations amènent des dilemmes psychologiques extrêmement profonds. Elles remettent en cause l'idée selon laquelle les comportements complexes des animaux dotés de cerveaux complexes peuvent être expliqués en faisant seulement appel à la notion d'« apprentissage ». En fait, cette crise met en évidence le problème fondamental des neurosciences : comment se fait-il qu'un animal, qui au départ n'a été confronté qu'à un petit nombre d'« événements » ou d'« objets », puisse ensuite classer par catégories, ou reconnaître, un nombre illimité d'objets nouveaux (même dans de nombreux contextes différents) semblables ou identiques au petit ensemble auquel il a été confronté initialement ? Comment se fait-il qu'un animal puisse reconnaître un objet sans l'aide d'un professeur ? Et comment se fait-il qu'il soit ensuite capable d'opérer une généralisation et de « construire des universaux » en l'absence de cet objet ou même en sa présence ? Ce genre de généralisation se passe du langage chez des animaux tels que le pigeon, comme nous le verrons plus loin.

Toutes les tentatives pour expliquer ces problèmes importants tendent soit à présupposer des indices cachés, qui ne seraient pas apparents pour l'expérimentateur, soit à traiter le monde de l'organisme étudié comme si ses « objets » et ses « événements » por-

taient des étiquettes. Mais en réalité, le monde, avec ses « objets », est dépourvu d'étiquettes et l'animal, pour obtenir des objets, définit les frontières macroscopiques de son environnement selon des modalités multiples et peut-être même infinies. Toutes les attributions de frontières par l'animal sont donc relatives et dépendent de ses besoins adaptatifs ou de ses intentions.

Le plus frappant, c'est que la capacité de découper le monde en « objets » et en agencements d'objets dépend du fonctionnement des cartes que nous avons mentionnées plus haut. Mais comment est-ce que les cartes interagissent pour définir des objets et des comportements tranchés ? Chez l'être humain, cette question conduit à ce que j'appelle la crise de l'homoncule. En effet, les processus perceptifs, dont on sait qu'ils sont fondés sur des sous-processus parallèles multiples et complexes, ainsi que sur de *nombreuses* cartes, ont un aspect unitaire pour celui qui les perçoit. (Dans le système visuel, il se pourrait qu'il existe plus de vingt centres cérébraux interconnectés possédant chacun sa propre carte.) Qui est l'organisateur de cette vision unitaire ? Résulte-t-elle de suites de « calculs », d'« algorithmes », qui se déroulent dans le cerveau – ou alors de l'action d'un homoncule, un petit bonhomme, qui à son tour a dans la tête un autre homoncule (voir figure 8-2), et ainsi de suite, à l'infini ? Qui se trouve là-haut ? S'il s'agit d'un homoncule, comment son cousin – celui que nous pourrions appeler l'électricien – aurait-il pu le construire, au cours du câblage qui s'opère durant le développement ? Comme nous l'avons vu, s'il existait effectivement, cet électricien mettrait en place un câblage pour le moins extrêmement curieux.

Où cela nous mène-t-il ? La réponse peut se résumer à ceci : « À un immense défi. » À moins que nous ne souhaitions faire des neurosciences de façon purement empirique, sans nous soucier le moins du monde de rechercher des explications cohérentes, nous devons affronter les crises que je viens d'évoquer. L'alternative évidente consiste à bâtir une théorie scientifique capable de réconcilier les contradictions et les dilemmes apparents, de résoudre les crises. Il est clair que, pour qu'une théorie du développement des fonctions cérébrales supérieures soit satisfaisante, elle doit se passer d'homoncules et d'électriciens à tous les niveaux. En même temps, elle doit rendre compte de la capacité à définir des objets et à les généraliser, laquelle opère sur un monde dont les événements et les « objets » ne sont pas étiquetés selon une procédure définie *a priori* ou un ordre hiérarchique. Cela ressemble bien peu aux tâches

que doivent accomplir les ordinateurs et beaucoup plus à des processus extrêmement inhabituels et très différents de ceux des ordinateurs.

Ce qui est particulier au cerveau – et qui fait défaut aux ordinateurs, aux particules matérielles, aux atomes, à la *res cogitans* et aux fantômes –, c'est le fait que sa morphologie est le produit d'une évolution. Comme nous l'avons vu, cette morphologie intervient à de nombreux niveaux, des atomes jusqu'aux muscles. La complexité des connexions cérébrales et leur nombre sont extraordinaires. Les cartes qui « se parlent » les unes aux autres sont parallèles et possèdent à la fois des caractéristiques statistiques et des caractéristiques précises. En outre, la matière de l'esprit interagit constamment avec elle-même. À ce propos, je n'ai pas encore évoqué ici le fait que les agencements dynamiques du cerveau sont des systèmes doués de mémoire ; autrement dit, les modifications antérieures de ces agencements influent sur leurs modifications ultérieures de façon spécifique et particulière. Le comportement du système nerveux est dans une certaine mesure auto-engendré par boucles : l'activité cérébrale conduit au mouvement, qui conduit à d'autres sensations et à d'autres perceptions, ce qui conduit encore à d'autres mouvements. Les niveaux et les boucles entre les niveaux constituent l'objet le plus complexe que nous connaissions, et ils sont dynamiques : ils se modifient constamment.

En fait, la dynamique chimique et électrique du cerveau ressemble davantage aux sons, aux lumières, aux mouvements et aux phénomènes de croissance qui agitent une forêt vierge qu'aux activités d'une entreprise productrice d'électricité. Cette dynamique résulte d'une chimie particulière. Toute modification de cette chimie, toute destruction de son substrat anatomique, peut provoquer des modifications mentales plus ou moins durables, qui vont de l'euphorie à la perte de connaissance et à la mort.

Tout en reconnaissant que la merveilleuse matière sous-jacente à l'esprit ne ressemble à aucune autre, nous devons nous méfier de tomber dans le chauvinisme primaire. Cette position consisterait à affirmer que *seules* les substances biochimiques dont le cerveau est constitué auraient été susceptibles d'aboutir à une telle structure. Or, même si cela était un tant soit peu le cas, il n'empêche que c'est l'organisation dynamique de ces substances, et non leur composition, qui engendre les processus mentaux – c'est cela qui est essentiel. C'est d'un bout à l'autre une affaire de morphologie dynamique. D'un bout à l'autre, oui, mais pas au point de nous obliger

à invoquer les événements et les forces physiques très particuliers mis en jeu entre particules physiques élémentaires. Pourtant, certains chercheurs, insuffisamment versés dans la morphologie du cerveau et dans les propriétés de la mémoire, ont été tentés d'expliquer les propriétés mentales à ce niveau – au niveau quantique [1].

S'il est exclu de faire preuve de chauvinisme biochimique, il n'est pas non plus permis de céder au libéralisme de l'informaticien : celui-ci part de l'idée d'un « logiciel » cérébral qui en fait n'existe pas *a priori*, pour ensuite affirmer que la structure sur laquelle on fait tourner ce logiciel n'a aucune importance. Ce faisant, il commet deux erreurs fondamentales : d'une part, aucun logiciel n'intervient dans les opérations du cerveau et, d'autre part, une masse écrasante de données incite à penser que la morphologie du cerveau intervient de façon écrasante.

Fort de ces précisions et d'un certain nombre de faits neuroscientifiques, nous pouvons à présent nous tourner vers des questions biologiques plus générales. Il est essentiel de les examiner si nous voulons éviter les écueils parsemant la voie qui nous amènera à mieux comprendre la matière de l'esprit.

1. Voir la postface.

DEUXIÈME PARTIE

Les origines

L'une des tentations auxquelles on s'expose quand on possède un esprit consiste à essayer de n'utiliser que lui pour résoudre le mystère de la nature qui lui est propre. Les philosophes s'y évertuent depuis la nuit des temps. Les psychologues cèdent également à cette tentation, comme nous tous, d'ailleurs, de temps à autre. Mais cette méthode ne permet pas d'explorer la matière de l'esprit.

Cependant, depuis que Darwin a suggéré que celui-ci résultait d'une évolution, nous disposons de moyens puissants pour comprendre comment il peut fonctionner. Cela signifie en effet qu'il n'a pas toujours existé ; il est apparu à un moment particulier, à l'issue d'une série d'étapes graduelles. Cela veut également dire que nous devons nous intéresser aux formes animales, parce que l'évolution nous enseigne que la sélection d'animaux capables d'assurer des fonctions qui améliorent leur adaptation au milieu ambiant est centrale pour la résolution du problème qui nous occupe.

Le « cerveau » du problème, c'est précisément le problème du cerveau, structure la plus complexe que nous connaissions dans l'univers. Pour la comprendre, nous devrons passer de la philosophie à l'embryologie, en effectuant un bond étrange mais nécessaire. Cela fait, nous serons en mesure de revenir à la philosophie via la biologie – ce qui nous occupera au cours des deux parties suivantes de ce livre.

Chapitre 4

Les bases biologiques de la psychologie

> *La psychologie constituait pour lui une nouvelle étude, et un coin obscur de l'éducation.* [...] *Il mit la psychologie sous clé ; il voulut que l'on respectât ses normes absolues ; que l'on tendît vers l'Unité ultime. La manie d'examiner tous les aspects de chaque question, de regarder par chaque fenêtre, d'ouvrir chaque porte, ôtait à ses femmes, comme le leur fit judicieusement remarquer Barbebleu, toute utilité pratique au sein de la société.*
>
> Henry Adams

Il est toujours risqué de négliger l'origine des choses. Encore plus quand il s'agit d'expliquer les phénomènes mentaux. Pourtant, c'est exactement ce qui s'est produit au cours de la majeure partie de l'histoire de la psychologie et de la philosophie de l'esprit. Je suppose que c'est dû au fait que la pensée est un processus réflexif et récursif. Par conséquent, il est tentant de penser qu'elle seule suffit à dévoiler sa propre nature. Mais lorsque nous revenons au précédent chapitre sur l'esprit, nous remarquons que la différence majeure entre les objets qui font preuve d'intentionnalité et les autres réside dans le fait que les premiers sont des entités biologiques. Cela ne signifie pas que l'intentionnalité est le propre de toutes les choses vivantes, mais simplement qu'elle n'est jamais le propre des objets inanimés. Comme je l'ai évoqué au dernier chapitre, nous devons rendre compte de la façon dont l'esprit s'incarne dans chaque individu.

Donc, nous devons tenir compte de la biologie. Mais pas seulement en raison de l'incarnation. Tout aussi importantes sont les données issues de l'évolution : elles incitent à penser que l'intentionnalité a émergé assez tardivement. Quelle est la base des phénomènes mentaux, et quand sont-ils apparus au cours de l'évolution ? Une réponse commode consiste à dire qu'ils ont émergé lorsque les animaux ont acquis un système nerveux. Toutefois, ce n'est pas tout à fait exact ; le simple fait de posséder des cellules nerveuses ne semble pas suffisant. Au cours de cette partie du livre, je compte m'intéresser à cette question des origines. Je voudrais démontrer

que la condition minimale nécessaire à l'apparition de phénomènes mentaux est un type spécifique de morphologie.

Mais auparavant, je tiens à montrer, plus généralement, qu'il est nécessaire de relier la psychologie à la biologie. C'est précisément parce qu'ils ont négligé toute considération biologique que les philosophes se sont engagés sur une fausse route. De même, une vision trop étroite de la psychologie peut également fourvoyer. Néanmoins, mon but n'est pas d'affirmer que, lorsqu'elles sont indépendantes de la biologie, les recherches philosophiques et psychologiques n'ont aucune valeur. Souvent, il a *fallu* qu'elles soient abordées en l'absence de données biologiques de base. Même une croyance erronée ou une fausse théorie sont susceptibles de stimuler une science et de lui donner vie jusqu'à ce que les données ou la méthodologie adéquates soient disponibles. Nous pouvons donc considérer ce chapitre comme un intermède historique, superficiel et sommaire, mais révélateur – du moins je l'espère – du riche écheveau de pensées qui ont alimenté le problème de l'esprit.

L'ignorance de la biologie lorsque l'on réfléchit à l'esprit et à l'acquisition des connaissances est de tradition. Pour une grande part, la philosophie de l'esprit a mené ses recherches sans se soucier (excepté de façon anecdotique) ni du corps ni du cerveau. Nous avons déjà vu que le premier philosophe moderne, Descartes, a fondé son rationalisme sur la pensée elle-même. Pour cela, il s'est appuyé sur son célèbre « doute méthodique », qu'il a esquissé dans son *Discours de la méthode* :

> Je pensai qu'il fallait [...] que je rejetasse, comme absolument faux, tout ce en quoi je pourrais imaginer le moindre doute, afin de voir s'il ne resterait point, après cela, quelque chose en ma créance, qui fût entièrement indubitable. Ainsi, à cause que nos sens nous trompent quelquefois, je voulus supposer qu'il n'y avait aucune chose qui fût telle qu'ils nous la font imaginer. Et parce qu'il y a des hommes qui se méprennent en raisonnant, même touchant les plus simples matières de géométrie, et y font des paralogismes, jugeant que j'étais sujet à faillir, autant qu'aucun autre, je rejetai comme fausses toutes les raisons que j'avais prises auparavant pour démonstrations [...] je me résolus de feindre que toutes les choses qui m'étaient jamais entrées en l'esprit n'étaient non plus vraies que les illusions de mes songes.

Descartes conclut qu'il existait une substance pensante. Ce faisant, il éluda radicalement la biologie ainsi que tout ce qui est d'ordre matériel. Étant donné les incursions remarquables de Des-

cartes en biologie, cela surprend. Il est cependant une question qu'il n'a pas analysée de manière explicite : le fait que pour être conscient et capable d'orienter sa pensée philosophique, le langage lui était nécessaire. Pour posséder un langage, il faut l'intervention d'au moins une autre personne, même s'il ne s'agit que du souvenir que l'on en garde – c'est-à-dire d'un interlocuteur intériorisé. Cette condition nécessaire ébranle l'idée cartésienne selon laquelle ses conclusions dépendraient de lui-même et de personne d'autre. De plus, Descartes ne précise pas à quel moment l'être humain devient une substance pensante. Peut-être aurait-il dû s'intéresser de plus près à la question de savoir s'il était vraisemblable qu'un bébé français en vienne à conclure « Je pense donc je suis ».

Les réponses d'ordre philosophique aux problèmes posés par le rationalisme cartésien, celles fournies par les empiristes britanniques John Locke, George Berkeley et David Hume par exemple, ne sont guère meilleures. L'idée de Locke selon laquelle l'esprit serait une ardoise vide, une *tabula rasa*, a été explorée sans référence aux données sur l'évolution et le développement qui permettent de penser que l'ensemble des répertoires de comportements se trouve sous contrôle génétique. Le monisme idéaliste de Berkeley – l'idée que, dans la mesure où toutes les connaissances sont acquises par l'intermédiaire des sens, le monde tout entier ne relève que de l'esprit – ne tient pas devant les données de l'évolution. Il serait effectivement très étrange que nous puissions créer un environnement qui, ensuite, serait capable de nous soumettre (mentalement) à la sélection naturelle.

Hume, le plus impitoyable et le plus sceptique des empiristes, conclut que, puisque toute connaissance est fondée sur des impressions sensorielles, la connaissance n'est jamais certaine. Même les connaissances scientifiques semblèrent ébranlées par son analyse des causes et des effets, lesquelles selon lui ne sont rien de plus que des corrélations mentales fondées sur la répétition des impressions sensorielles. Mais, comme nous le verrons plus loin, la question ne se limite pas aux impressions sensorielles ; la biologie de l'esprit est bien plus étendue.

Emmanuel Kant, plus versé en physique et en astronomie qu'en biologie, a situé le problème dans une perspective plus large. Répondant à Hume, il a fait remarquer qu'il existe, dans l'esprit, des catégories *a priori*, dont l'harmonie avec l'expérience sensible est par là même garantie. L'existence de catégories *a priori* semble mieux s'accorder avec les données actuelles sur les comporte-

ments éthologiquement déterminés et sur les propriétés neurophysiologiques des cellules cérébrales. Pourtant, elle n'est pas totalement compatible avec ce que l'on sait sur la façon dont les bébés acquièrent le sens de l'espace au cours de leur développement – ni même avec la physique relativiste. Néanmoins, étant donné son inévitable ignorance des progrès modernes de la biologie et de la physique, on ne peut que pardonner à Kant de ne pas avoir compris quelles étaient les contraintes pouvant peser sur la notion d'*a priori*.

Je pourrais citer d'autres exemples encore. Ceux que je viens d'évoquer montrent qu'en philosophie une connaissance de la psychologie fondée sur l'expérimentation, et une compréhension de la neurologie et de l'évolution, sont utiles à qui veut éviter les erreurs grossières. Cependant, toutes ces connaissances n'ont été acquises que récemment, et on ne peut qu'admirer le courage et la persévérance de ces grands penseurs, qui ont approfondi toujours plus ces problèmes si importants.

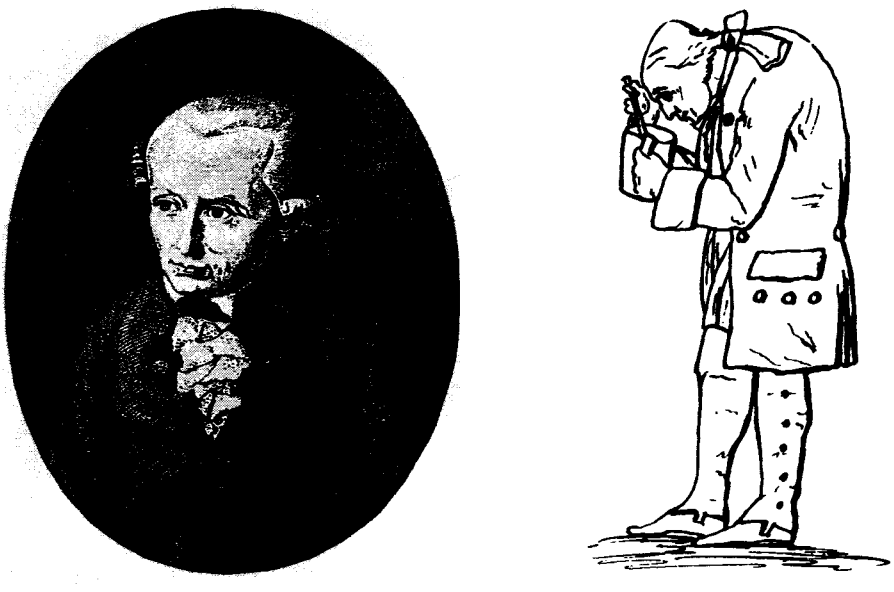


FIGURE 4-1

Emmanuel Kant (1724-1804), le grand philosophe du Siècle des Lumières. Ses idées transformèrent profondément les visions rationaliste et empiriste de l'esprit. Le dessin représente le grand homme en train de préparer de la moutarde.

En l'absence de connaissances sur le cerveau et le système nerveux, la psychologie elle-même ne s'en est pas non plus très bien

sortie. Pour autant, les psychologues ont accumulé une énorme quantité d'informations utiles et importantes depuis que William James, à Harvard en 1878, et Wilhelm Wundt, à Leipzig en 1879, ont fondé les premiers laboratoires de psychologie physiologique expérimentale. Néanmoins, au lieu d'aboutir à une théorie unifiée de l'esprit, ce mouvement a donné lieu à toute une série d'écoles, et chacune a développé une conception différente des comportements, de la conscience et de l'importance relative de la perception, de la mémoire, du langage et de la pensée.

Il ne m'appartient pas ici d'examiner dans le détail le point de vue de chacune de ces écoles. Mais il peut être utile de mentionner quelques-unes de leurs principales lignes de pensée afin de mettre en évidence la nécessité d'un dénominateur commun biologique. James lui-même fut l'un des plus grands pionniers de la psychologie moderne. Dans ses *Principles of Psychology*, il arguait que, tout en prêtant attention au cerveau, la psychologie pouvait procéder à l'étude des fonctions mentales de façon indépendante : il lui suffirait de combiner adéquatement introspection, expériences et psychophysique. À la même époque, en Allemagne, la psychophysique connut également des progrès grâce aux travaux de Wilhelm Wundt, de Ewald Hering et du grand physicien Hermann von Helmholtz. Ces travaux consistaient à mesurer très soigneusement les temps de réponse et les jugements suscités par des stimulus physiques eux-mêmes mesurés avec précision.

James vaut peut-être surtout pour avoir remarqué, lorsqu'il a caractérisé ce mystérieux processus dans son essai intitulé « La conscience existe-t-elle ? » puis dans les *Principles*, que la conscience est un processus et non une substance. Selon Whitehead, James fut pour le XXe siècle ce que Descartes fut pour le XVIIe. Du temps de James, cependant, on utilisait encore trop souvent l'introspection pour parvenir à des conclusions concernant l'esprit, souvent avec des résultats douteux. C'était le cas d'Edward Titchener, qui considérait l'introspection expérimentale comme « l'unique porte d'accès à la psychologie » et élaborait de grandes théories des sensations et des sentiments fondées sur cette méthode. De même, les études sur la mémoire humaine (par exemple, celles d'Hermann Ebbinghaus) utilisaient des séquences et des syllabes abstraites, dénuées de sens. Elles ne s'intéressaient quasiment pas au rôle du sens dans la mémoire.

Les expériences réalisées au début du XXe siècle par Ivan Pavlov sur les réflexes conditionnés prirent le contre-pied de ces approches.

Après avoir reçu un stimulus inconditionnel (de la nourriture), accompagné à maintes reprises d'un stimulus conditionnel (une sonnerie), des animaux salivent lorsque, par la suite, on leur fait entendre uniquement le signal sonore. Edward Thorndike et Clark Leonard Hull, aux États-Unis, ont généralisé et analysé de façon plus approfondie les paradigmes stimulus-réponse. Au bout d'un certain temps, l'idée extrémiste selon laquelle la seule étude scientifique possible en psychologie était l'étude des comportements est apparue. Comme l'a dit John Watson, le behaviorisme rendait hors la loi la conscience, l'introspection et autres. Le plus féroce partisan de cette position fut par la suite B. F. Skinner, qui étudia longuement le phénomène dit de conditionnement opérant. (Au lieu de réagir à un stimulus conditionnel classique, l'animal est récompensé au cours d'un comportement opérant particulier. Ce comportement est ensuite renforcé par des récompenses répétées.)

De nombreux enchaînements complexes de comportements ont été analysés au moyen des techniques behavioristes. Il est clair, cependant, qu'au moins une partie du bébé partait à chaque fois avec l'eau du bain. Par exemple, cette approche ne tenait pas compte des phénomènes de forme *(Gestalt)* (figure 4-2) découverts par Max Wertheimer, Wolfgang Köhler et Kurt Koffka. De fait, le behaviorisme avait du mal à rendre compte de la façon dont les sujets pensants sont capables de discerner des formes. La conscience est impossible à chasser. Les observations de Sigmund Freud, qui avait remarqué les effets du refoulement sur la mémoire et ceux de l'inconscient sur les comportements conscients, mettaient en évidence les failles de l'explication behavioriste. Par ailleurs, les expériences de Sir Frederic Bartlett sur la mémoire humaine incitaient à penser que la mémoire était plus qu'une simple répétition machinale de séquences de caractères dépourvus de sens, comme avaient paru l'impliquer auparavant les travaux d'Ebbinghaus. La biologie et la nature humaine revenaient au galop sur des questions que le behaviorisme avait négligées.

L'un des aspects importants qu'il fallait expliquer à propos de la nature humaine et du comportement avait été mis en évidence par la médecine clinique. En effet, au cours du XIXe siècle, Gustav Fritsch et Julius Hitzig avaient découvert, d'une part, l'existence de cartes cérébrales en constatant que lorsqu'ils stimulaient électriquement certaines parties du cerveau de leurs patients, ils induisaient chez eux des mouvements corporels spécifiques. D'autre

part, Paul Broca avait découvert que des lésions localisées dans certaines régions de la partie gauche du cerveau induisaient des aphasies motrices (l'incapacité de produire un discours cohérent). Ni l'une ni l'autre de ces découvertes ne pouvait être ignorée. Rapidement, les écoles de neurophysiologie foisonnèrent et, dès la fin du XIX[e] siècle, les chercheurs étaient déjà bien engagés sur la voie qui les mènerait à mesurer effectivement l'activité neuronale. Entre les deux grandes guerres, grâce à une série d'innovations techniques mises au point par Sir Charles Sherrington, il devint possible de détecter à la fois l'activité individuelle et l'activité collective des cellules nerveuses.

La psychologie présentait donc une image très hétéroclite : behaviorisme, psychologie gestaltiste, psychophysique et étude de la mémoire en psychologie normale ; études des névroses en analyse freudienne ; études cliniques des lésions cérébrales et des défauts sensoriels et moteurs, classification des psychoses et de leurs déroutants symptômes, en médecine ; et, en physiologie, accumulation d'une masse croissante de connaissances concernant à la fois la neuro-anatomie, grâce aux travaux de pionnier de Santiago Ramón y Cajal, et le comportement électrique des cellules nerveuses (grâce à Sherrington). Mais ce n'est qu'occasionnellement que des chercheurs – tels que Karl Lashley ou Donald Hebb – s'efforcèrent sérieusement de relier, à un niveau plus général, ces domaines disparates. La plupart du temps, chaque discipline se développait indépendamment des autres, et les chercheurs s'appliquaient souvent à réfuter les idées défendues par leurs concurrents des disciplines « extérieures ».

Ce qui surprend, c'est la relative indépendance de ces différentes voies par rapport à la théorie de l'évolution, laquelle est absolument indispensable à quiconque veut comprendre l'esprit. Darwin a énoncé la théorie de la sélection naturelle en 1859. Pour lui, il était clair que l'évolution affectait nécessairement les comportements et vice versa. Mais il n'y eut que deux de ses contemporains – George Romanes et C. Lloyd Morgan – pour prôner l'idée selon laquelle il existait un lien entre l'évolution et le comportement. Quant aux effets du développement sur les comportements, C. W. Mills et J. M. Baldwin en ont vu l'importance, mais leurs idées, qui ont contribué à fonder notre vision moderne, sont restées marginales. Ce n'est que plus tard, avec les études réalisées par Jean Piaget sur le développement des comportements cognitifs chez l'enfant, qu'ont

été jetées les bases qui allaient servir aux études modernes du développement des capacités cognitives.

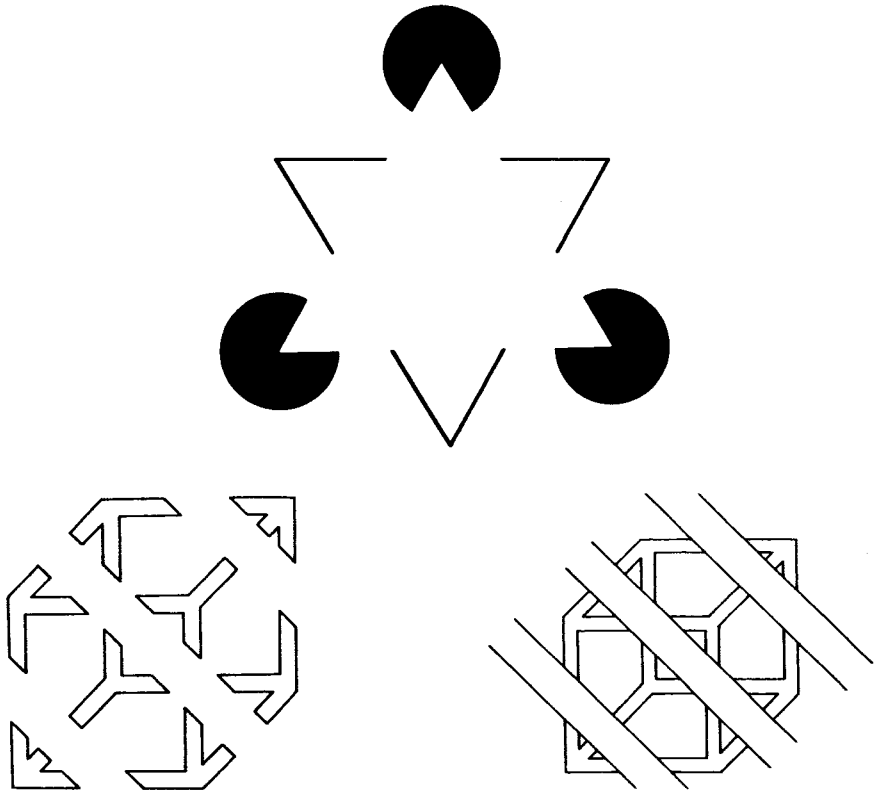


FIGURE 4-2

Phénomènes de forme. Ces figures, extraites des travaux de Gaetano Kanizsa, montrent à quel point la perception dépend du contexte. (Cachez les ronds noirs de la figure du haut et voyez les contours apparents disparaître.) Comme l'a dit Kanizsa : « Voir et penser sont des activités clairement distinctes. Avec ces " morceaux ", nous pouvons imaginer *un cube* (en bas à gauche), *mais il nous est très difficile de le voir. » Remarquez, cependant, qu'il suffit de superposer trois barres opaques* (en bas à droite) *pour compléter le cube, qui devient alors* présent *du point de vue perceptif.*

Bien sûr, des théories diamétralement opposées ont également été élaborées : depuis les travaux d'Herbert Spencer, autre contemporain de Darwin, on a constamment essayé d'expliquer les comportements en termes de psychologie sociale. Ces tentatives en sont généralement restées à un niveau descriptif et ont toujours reposé sur des caractéristiques culturelles et sur ce que l'on appelle la « psychologie populaire » – c'est-à-dire l'évaluation des comportements humains au moyen du sens commun. Les controverses et

les spéculations abondent sur l'inné et l'acquis, sur la contribution de la génétique et celle de l'environnement. La richesse et l'étendue de cet ensemble d'études ne se prêtent pas facilement à la synthèse. Néanmoins, à mesure que les techniques modernes permettant de mesurer la fonction cérébrale se sont développées et que notre compréhension de la biochimie du cerveau s'est améliorée, il est devenu clair qu'on ne peut faire de la psychologie qu'en l'enracinant de plus en plus dans la biologie. Au mieux, on peut provisoirement continuer à l'étudier comme par le passé, en attendant les interprétations biologiques.

Une fois parvenu à cette conclusion, on ne peut échapper à une autre, encore plus fondamentale : les phénomènes psychologiques dépendent de l'espèce dans laquelle on les observe, et les caractéristiques des espèces dépendent de la sélection naturelle. Ce point de vue, adopté par des éthologues comme Nikolaas Tinbergen et Konrad Lorenz, ainsi que par la plupart des psychologues modernes, relie inévitablement la psychologie à la biologie. Et l'existence même d'une telle liaison constitue une preuve de l'importance de l'évolution dans le comportement des espèces.

Lorsque nous nous intéressons à l'esprit, nous devons tenir compte à la fois de nos similitudes et de nos différences au regard des autres espèces. Comme nous le verrons au chapitre 16, l'une de ces différences tient au fait que chacun de nous possède une « âme » individuelle fondée sur le langage. Malheureusement, quoi que nous réussissions à découvrir sur les propriétés du langage, ni la psychologie ni la biologie ne permettront la transmigration des âmes. Cela me rappelle l'histoire d'un mourant qui, pour consoler sa femme éplorée, lui promit qu'il reviendrait exactement six semaines après sa mort, et lui demanda d'aller consulter un médium le moment venu. Réconfortée, la veuve attendit patiemment, puis alla voir le médium, comme convenu. Une voix, sortie d'un recoin sombre de la pièce, lui dit alors : « Bonjour, ma chérie. – Harry, dit-elle, est-ce bien toi ? » Et la voix lui répondit : « Bien sûr que c'est moi. » Elle demanda alors prudemment : « Que fais-tu de tes journées ? » Et la voix lui répondit : « Je me lève, je fais l'amour, je me promène, je fais l'amour, je déjeune, je fais l'amour, je fais la sieste, je fais l'amour. Et le lendemain, ça recommence. » La femme digéra lentement ces paroles et dit avec une certaine appréhension : « Mais chéri, je ne savais pas que les anges du ciel faisaient l'amour. » Et la voix répondit : « Je ne suis pas un ange et je ne suis pas au ciel. Je suis un lapin et j'habite le Saskatchewan. »

Bien qu'il soit nécessaire de tenir compte des idées des philosophes et de celles des diverses écoles de psychologie dès que l'on s'intéresse à la question de l'esprit, ce n'est que récemment que ces idées ont été confrontées aux questions clés de la biologie elle-même. Le message se résume à ceci : la morphologie (l'anatomie) animale et celle des espèces, ainsi que la façon dont cette morphologie fonctionne, sont les bases premières de tout comportement et de l'émergence de l'esprit. La sélection naturelle agit sur les individus au cours de leur compétition intra- et inter-spécifique. Et, d'après l'étude des données paléontologiques, on sait que ce que nous appelons l'esprit n'est apparu qu'à des moments particuliers au cours de l'évolution (et plutôt tardivement, d'ailleurs).

Ces quelques remarques austères peuvent constituer le point de départ d'un programme de recherche visant à relier la psychologie à la biologie et à rendre compte de la façon dont s'incarne l'esprit. Étant donné ce que nous savons sur l'histoire de la philosophie de l'esprit et de la psychologie, il est peu vraisemblable que le fait de continuer à éluder les aspects biologiques fondamentaux d'un tel programme nous aide à mieux comprendre comment l'esprit humain est apparu et comment il fonctionne. Dans la postface, je passe en revue les erreurs qui subsistent lorsque l'on s'obstine à pratiquer la psychologie sans se référer à la biologie.

Au centre de toute liaison entre ces deux domaines on trouve, bien entendu, les données de l'évolution. Darwin a été le premier à reconnaître que la sélection naturelle devait rendre compte même de l'émergence de la conscience humaine. Examinons à présent certaines de ses idées et les conséquences qui en découlent.

Chapitre 5

L'esprit et sa morphologie ou comment achever le programme de Darwin

> *Mais voici que survient le doute : peut-on faire confiance à l'esprit de l'homme, qui s'est développé, j'en suis persuadé, à partir d'un esprit aussi primitif que celui des animaux les plus primitifs, lorsqu'il en vient à tirer des conclusions aussi grandioses ?*
>
> Charles Darwin

Alfred Wallace, co-découvreur de la théorie de la sélection naturelle, écrivit une série de lettres à Charles Darwin, pour lui exprimer ce qu'il considérait comme un point de vue hérétique. Wallace rejetait le fait que la sélection naturelle puisse rendre compte de l'évolution humaine. Selon lui, les capacités de l'esprit humain ne pouvaient pas s'expliquer par la seule sélection naturelle.

Darwin adopta la position inverse. Il ne voyait aucune raison de penser que la sélection naturelle ne puisse pas donner lieu aux caractéristiques de base sous-jacentes à la pensée humaine. Il consacra à cette idée deux de ses livres, *La Descendance de l'homme et la sélection sexuelle* et *L'Expression des émotions chez l'homme et les animaux.*

Il est important de comprendre les idées de Darwin sur l'évolution et la sélection naturelle. Énoncées simplement, elles stipulent que l'évolution est le résultat à la fois de la compétition entre individus et des modifications de l'environnement, qui agissent toutes deux sur les fluctuations internes des populations (figure 5-2). Il existe toujours une variabilité au sein d'une population d'êtres vivants, et celle-ci donne lieu à des différences d'aptitudes. La sélection naturelle induit alors la reproduction différentielle des individus dont les variations (lire compétences structurelles et fonctionnelles, ou phénotype) les dotent, ainsi que leur progéniture, d'avantages statistiques leur permettant de s'adapter aux modifications de l'environnement ou de concurrencer les autres individus de la même espèce ou d'une autre. Ainsi, la reproduction différentielle et l'hé-

rédité augmentent la probabilité que les caractéristiques qui améliorent l'adaptation soient préservées.

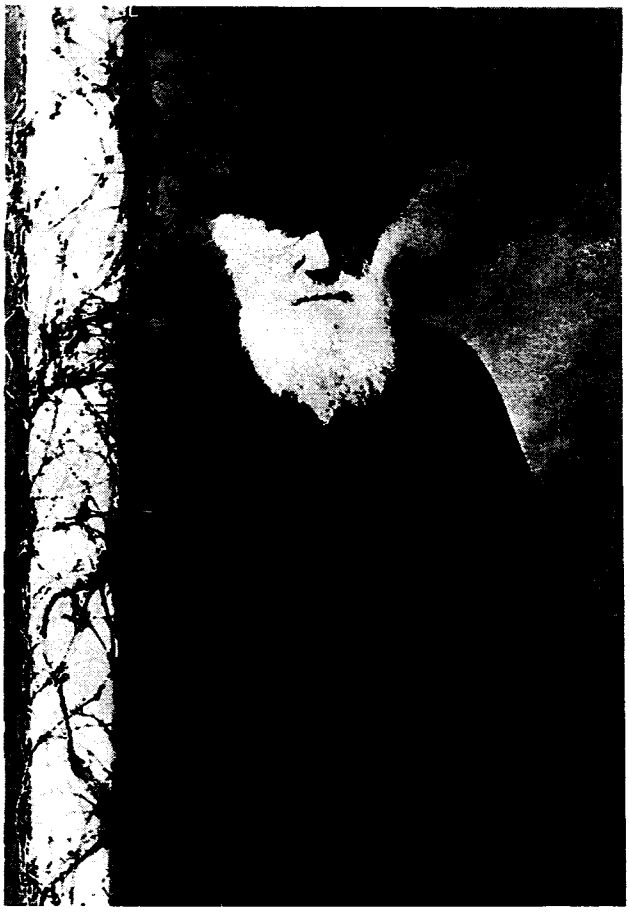
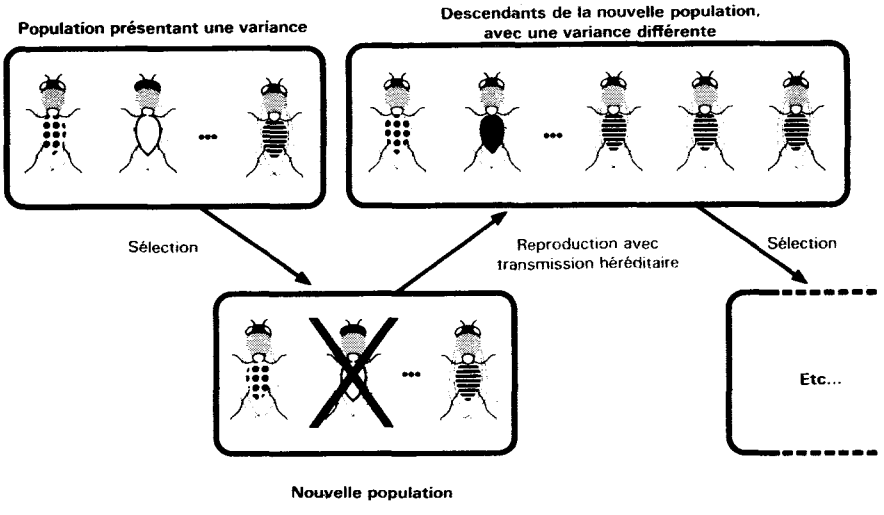


FIGURE 5-1

Charles Darwin (1809-1882), fondateur de la théorie moderne de l'évolution – la base théorique de l'ensemble de la biologie. Darwin affirma que l'évolution des êtres humains en tant qu'espèce était sujette aux mêmes types de forces qui conduisent à l'évolution des autres espèces. Il devint assez renfermé à la fin de sa vie, comme le montre, sur cette photographie, son air digne mais lugubre.

Le changement fondamental induit dans la population résultante concerne la fréquence des gènes responsables de l'existence de ces caractères avantageux. (J'aborderai plus longuement la question des gènes et de la génétique au cours du chapitre suivant.) La modification des fréquences génétiques est la preuve du *fait* qu'une

évolution a bien eu lieu. Mais c'est la sélection naturelle des phénotypes (les capacités structurelles et fonctionnelles globales) des individus qui constitue *le moyen* par lequel cette évolution s'opère. Et le principal *niveau* auquel se produit la sélection est celui de l'individu et de son comportement. Ce qu'il nous faut donc comprendre, ce sont les règles qui relient la sélection et l'expression des gènes aux diverses modifications que les gènes induisent sur le phénotype (figure 5-3). Il s'agit là d'une tâche gigantesque, qui n'a été que partiellement menée à bien jusqu'ici.



PENSER EN TERMES DE POPULATIONS

FIGURE 5-2

Les variations sont dues aux mutations survenant au sein d'une population d'organismes. La sélection naturelle favorise la reproduction différentielle des membres de la population qui sont en moyenne les plus aptes. Il en résulte une augmentation dans la population de la fréquence relative des gènes qui confèrent un avantage adaptatif. Remarquez la sélection au détriment *des yeux sombres et* en faveur *des corps rayés.*

Bien que Darwin n'ait pas correctement perçu les mécanismes génétiques en jeu, il ne s'est pas trompé sur le principe de base. Il a compris, par exemple, qu'il existe vraisemblablement des similitudes phénotypiques dans l'expression des affects entre certains animaux et les êtres humains. Il a également saisi que la sélection naturelle n'a pas nécessairement sélectionné *directement* toutes les expressions affectives. Le même raisonnement vaut pour la pensée et le comportement. Le point de vue de Darwin consistait à dire

que les modifications survenant graduellement au sein des populations pouvaient même rendre compte de l'apparition et de l'origine des êtres humains.

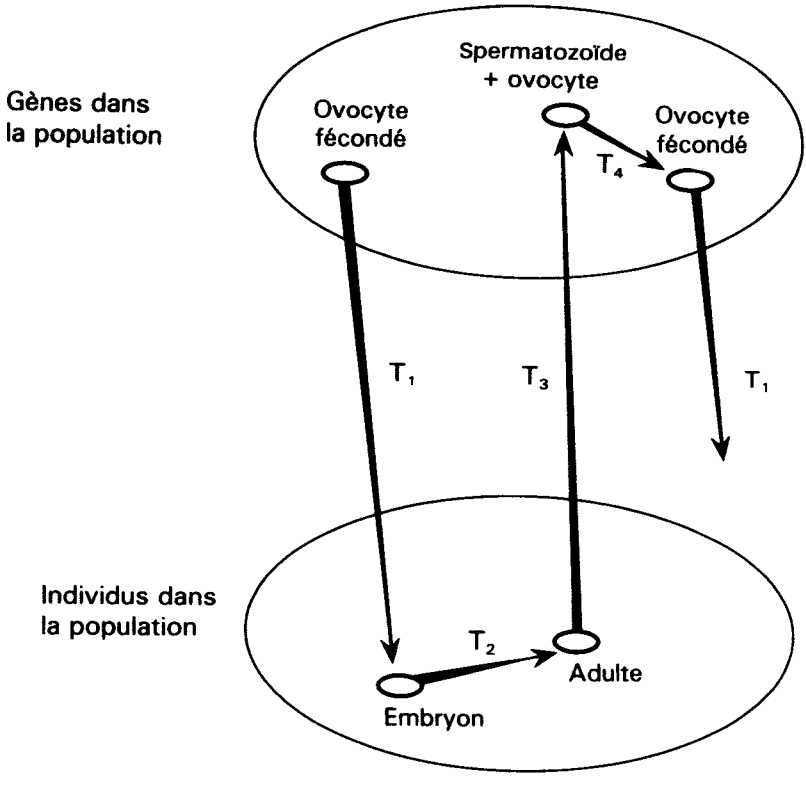


FIGURE 5-3

Les modifications de la fréquence des gènes peuvent être reliées au processus de sélection naturelle lui-même. L'augmentation relative, dans une population donnée, des gènes ayant amélioré l'adaptation au milieu, est à la fois le résultat *et le* signe *d'une évolution. Mais la sélection s'exerce sur des* individus, *que ce soit sur les spermatozoïdes ou les ovocytes, sur l'embryon, ou encore sur l'animal dans son environnement. Par conséquent, nous devons élucider les règles reliant le développement et le comportement aux effets des gènes [les lignes verticales pour les transformations* (T_1-T_4)]. *T_1 représente le développement ; T_2 représente le passage à la vie adulte dans le milieu naturel ; T_3 représente la formation des spermatozoïdes et des ovocytes ; T_4 représente l'ovule fécondé, prêt à entamer un nouveau cycle de développement.*

Mais cette idée était en grande partie indémontrable du temps de Darwin. Un grand nombre d'éléments lui échappaient encore – comme, par exemple, la véritable nature de l'hérédité génétique, les données essentielles sur les restes fossiles des hominidés, et une grande quantité d'informations importantes sur la façon dont se

développent les animaux. Néanmoins, il se trouve que sa démarche de base était correcte. Elle consistait à déterminer ce qu'il nous faut savoir pour rendre compte de l'origine de l'esprit humain au cours de l'évolution – j'appelle cela le programme de Darwin. Ce que nous devons parvenir à comprendre (hormis les mécanismes de l'hérédité), c'est comment la morphologie sous-jacente au comportement a émergé au cours de l'évolution des espèces, et comment le comportement lui-même modifie la sélection naturelle. Si j'appelle cela le programme de Darwin, ce n'est pas parce qu'il représente tout ce que Darwin souhaitait savoir, mais parce que c'est ce qui l'a le plus intéressé durant la dernière partie de sa vie. Bien sûr, Darwin n'a pas réalisé son programme. Mais si nous acceptons comme lui l'idée que tous les aspects du comportement humain finissent par s'expliquer en termes évolutifs, alors nous devons essayer d'achever ce programme.

De quoi avons-nous besoin pour ce faire ? Premièrement, d'analyser les effets de l'hérédité sur le comportement et vice versa. Deuxièmement, de déterminer comment la sélection naturelle influe sur le comportement et est à son tour influencée par lui. Troisièmement, de comprendre comment la forme animale – ou morphologie – rend le comportement possible et quelles sont les contraintes qu'elle lui impose. Quatrièmement (et surtout), de comprendre comment la morphologie animale apparaît et se modifie au cours du développement. (Par morphologie, je n'entends pas simplement la forme – les extrémités, la symétrie, etc. –, mais aussi les détails microscopiques qui confèrent leurs fonctions aux tissus et aux organes tels que le cerveau.) Cette dernière condition requiert que l'on sache relier évolution et morphogenèse au cours du développement. La nature de ce lien constitue l'énigme la plus importante et la plus centrale de la biologie moderne – l'énigme de l'évolution morphologique.

Il peut être utile, à présent, d'énoncer ce que l'on sait à propos des autres conditions du programme de Darwin. L'une d'elles a déjà été relativement satisfaite : nous savons que les bases de l'hérédité résident dans les gènes et nous comprenons pas mal de choses sur la façon dont ceux-ci sont transmis, modifiés et exprimés (figure 5-4). La collaboration entre généticiens et spécialistes de l'évolution, au cours des années 1940, a permis de relier les découvertes de Gregor Mendel en génétique à la théorie de l'évolution d'une manière extrêmement fructueuse. Cette « synthèse moderne » permit d'établir, d'une part, que les mutations survenant au sein

de l'acide désoxyribonucléique (ou ADN) sont à l'origine de la variabilité génétique (ce que Darwin n'avait pas pu faire) et, d'autre part, que les structures génétiques se réorganisent au cours d'un processus appelé recombinaison. Bref, on a commencé à identifier effectivement les résultats de la sélection naturelle en termes de modifications de la fréquence des gènes dans les populations.

Les découvertes ultérieures concernant la nature de l'ADN, ainsi que la possibilité de manipuler cette molécule au point d'insérer des gènes étrangers dans des animaux afin de modifier leur morphologie ou leur comportement, ont constitué une confirmation triomphale du point de vue adopté dans la synthèse moderne. De plus, d'énormes progrès ont été accomplis qui ont permis de généraliser les idées de Darwin, selon lesquelles l'émergence des différentes espèces était due à l'isolement sexuel ou géographique de groupes d'individus fertiles.

Suite à la synthèse moderne, un certain nombre de chercheurs ont commencé à étudier les comportements du point de vue de la génétique, de l'évolution et des interactions entre espèces. Cela marqua les débuts de l'éthologie, science dont les résultats prêchent en faveur de l'idée que certains profils de comportement varient selon les espèces et qu'ils sont donc sujets à l'influence génétique. Les découvertes des éthologues sont cependant plus subtiles. Elles incitent en effet à penser que des comportements complexes tels que le chant des oiseaux ont à la fois des composantes génétiques *et* épigénétiques. Par exemple, chez des espèces comme les passereaux, certains aspects du comportement moteur sous-jacent au chant sont donnés à la naissance, dans le phénotype, de même que certaines variantes et certaines modifications de la structure des vocalisations. Mais, pour être en mesure de chanter la chanson caractéristique de l'espèce de passereaux correspondant à une région donnée, ces oiseaux ont besoin d'entendre les chansons des oiseaux adultes de la même espèce. Chez des espèces comme celles-ci, les oiseaux rendus sourds à la naissance n'acquièrent jamais totalement le chant caractéristique de leur espèce. Il aurait fallu pour cela que des événements épigénétiques, faisant intervenir des interactions avec d'autres oiseaux, aient lieu.

Il n'est pas très difficile d'imaginer comment les différents types de comportement pourraient affecter et être affectés par la variabilité génétique et la sélection naturelle. Des études de ce genre, qui relient les fonctions des diverses régions du cerveau au reste du phénotype, contribuent à l'achèvement de certaines parties du

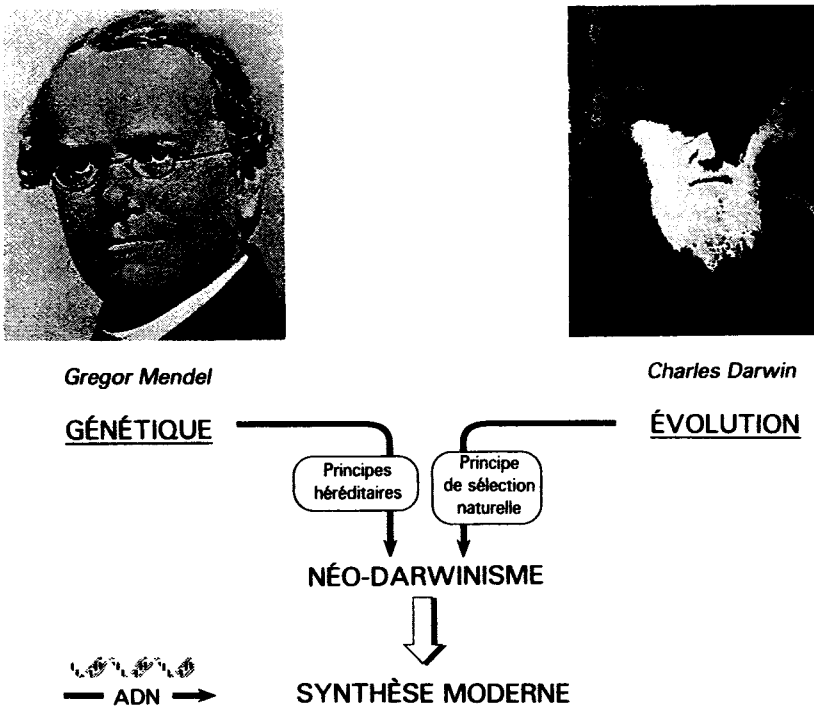


FIGURE 5-4

La synthèse moderne. Dans les années 1940, un groupe d'évolutionnistes et de généticiens réussit à concilier les découvertes initiales de Gregor Mendel (1822-1884) sur l'hérédité et la théorie de l'évolution décrite par Darwin (1809-1882). La théorie initiale de Darwin faisait intervenir une notion erronée d'hérédité. Mendel, un moine augustin autrichien, posa les bases de la génétique moderne, mais sa contribution ne fut pas reconnue tout de suite. Il fallut attendre l'année 1901 pour que ses résultats soient redécouverts.

programme de Darwin. Mais on ne doit pas abuser de cette façon de procéder. L'évolution de la sociobiologie nous met d'ailleurs en garde contre cela. Les sociobiologistes s'évertuent à expliquer les comportements d'après la sélection naturelle, et en particulier l'altruisme – ce qui illustre bien mon propos ici. En effet, si la sélection naturelle tend à maximiser les aptitudes des individus, on voit mal comment les gènes des individus qui se sacrifient avant de se reproduire, ou qui perdent une partie de leur aptitude reproductrice au service des autres, pourraient être transmis. Néanmoins, l'analyse génétique des abeilles soumises à ce qu'on appelle la sélection de parentèle montre que les femelles qui se trouvent au service d'une

reine sœur peuvent, au détriment de leur propre reproduction, faire augmenter la fréquence de leurs gènes dans la population. Ce résultat dépend des caractéristiques génétiques tout à fait particulières des abeilles et constitue une élégante victoire expérimentale. Mais les tentatives d'explication de l'altruisme humain en tant que conséquence directe de l'existence de « gènes de l'altruisme » sont une autre histoire – peu crédible, d'ailleurs.

Les gènes ne modifient pas directement la morphologie, mais agissent plutôt par combinaisons complexes. Et la morphologie modifie le comportement de manière subtile. Plus révélateur encore est le fait que de subtiles modifications morphologiques peuvent parfois induire des modifications tout à fait extraordinaires des comportements. Ce que nous voudrions savoir, c'est comment les modifications morphologiques – soit au niveau de l'animal tout entier, soit aux niveaux microscopiques du cerveau, des muscles ou des os – affectent les comportements, et comment les comportements affectent la morphologie. C'est cette partie du programme de Darwin qui se trouve encore largement inachevée.

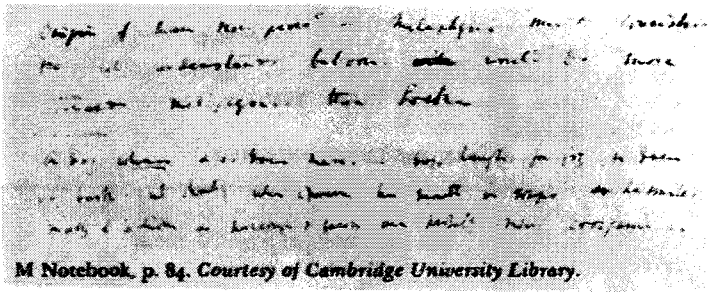
Lorsque l'on étudie de près les journaux de Darwin, on s'aperçoit que c'était un homme extraordinaire. Dans l'un de ces journaux, le cahier M, on peut lire : « Origine de l'homme démontrée. – La Métaphysique doit fleurir. – Celui qui comprend le babouin en fera plus que Locke pour la métaphysique. » Durant les dernières décennies de sa vie, triste et plus ou moins reclus, Darwin poursuivit infatigablement ses recherches. Mais ce n'est que maintenant que l'on commence à se rendre compte de tout ce qu'il savait. Ses réflexions sur les comportements étaient aussi profondes que ses réflexions sur la morphologie. C'était un grand penseur, mais jamais il ne lui serait venu à l'esprit d'essayer d'avoir l'air « futé ».

À ce propos, il est impossible d'échapper au problème posé par l'idée que les gènes spécifient des comportements complexes directement en essayant d'invoquer la « sélection des groupes ». Selon cette idée, par exemple, la sélection favoriserait les troupeaux rapides plutôt que les individus rapides au sein des troupeaux. Darwin envisagea et examina une telle possibilité. Cependant, hormis de rares exceptions, il semble que la plupart du temps la sélection naturelle se produise non pas au niveau des gènes ou des groupes d'individus, mais plutôt au niveau des individus eux-mêmes.

Ces considérations ne font encore une fois que mettre l'accent sur la partie du programme de Darwin qu'il est le plus urgent

83e La possibilité de deux suites tout à fait distinctes se déroulant dans l'esprit comme dans le dédoublement de la conscience explique peut-être vraiment ce qu'est une habitude – Dans la suite *habituelle* de pensées, chaque idée en *appelle d'autres, & la conscience de double individu n'est pas éveillée.* – L'individu habituel se souvient des choses faites dans l'autre état habituel parce que celui-ci modifie (sans qu'il en soit directement conscient ?) ses habitudes. 16 août. Exemple d'hérédité mentale. Suis un Darwin & tiens de mon Père par *principe héraldique.* & Eras un Wedgwood à beaucoup d'égards & un peu de Tante Sarah. excentriques [1], tout comme Catherine à certains égards – bons exemples. – quand l'éducation est la même. – Mon écriture la même que celle de Grand-Père. [2]

84e *16 août* Colère « Rage » sous la pire des formes est décrite par Spenser (Reine des Fées. *CD 25* (Description Reine) « O » de l'Enfer Chant IV ou V.) comme *pâle & tremblante.* & non comme rouge & les muscles rigides. – [1] Comment cela se fait-il ? *traité p. 241* [2] Origine de l'homme désormais démontrée. – La Métaphysique doit fleurir. – Celui qui comprend le babouin (en fera) en ferait plus que Locke pour la métaphysique



M Notebook, p. 84. *Courtesy of Cambridge University Library.*

FIGURE 5-5

Extrait des cahiers de Darwin.

d'achever. Elle concerne le problème de l'origine ancestrale de la morphologie animale, de la structure et de la fonction des tissus – c'est-à-dire de l'évolution morphologique. Pour comprendre pourquoi ce problème est si important, il suffit de songer aux extraordinaires données fossiles sur les hominidés, qui portent à croire qu'une grande augmentation de la capacité crânienne et de la taille du cerveau s'est produite chez les hominidés en moins d'un million d'années d'évolution (voir la figure 5-6).

Comment cette augmentation a-t-elle pu se produire si rapidement ? Quel rapport y a-t-il entre cette augmentation et les autres caractéristiques des hominidés dont nous pouvons déduire l'existence d'après les données fossiles et archéologiques ? Comment relier la morphologie globale et le comportement à la morphologie microscopique du cerveau ? Comment ces événements évolutifs s'arti-

culent-ils avec le comportement des hominidés en groupe et avec le développement du langage ?

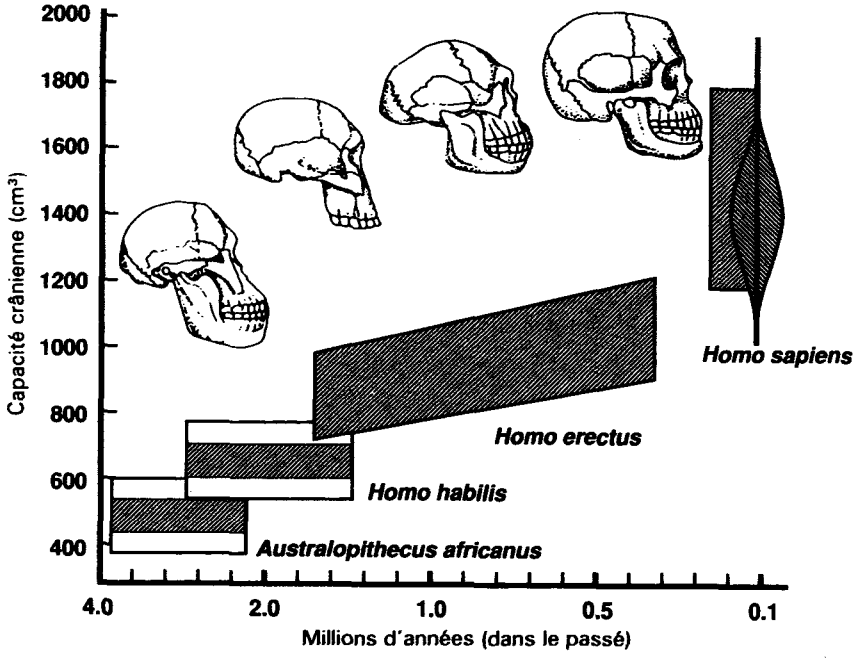


FIGURE 5-6

L'augmentation remarquable de la capacité crânienne au cours de deux millions d'années d'évolution de l'espèce humaine. Nous sommes des Homo sapiens sapiens *; les autres sont nos présumés ancêtres. Les quadrilatères hachurés représentent la période pendant laquelle vécut chaque espèce et la taille du cerveau au sein de chacune d'entre elles. Le rectangle correspondant à l'*Homo sapiens *comprend l'*Homo sapiens sapiens *et l'homme de Néanderthal. Le fuseau hachuré représente la distribution des capacités crâniennes chez* l'Homo sapiens sapiens *moderne. Au départ, ces crânes n'étaient pas vides ! Une théorie adéquate du cerveau doit être capable de rendre compte de cette énorme augmentation de la taille du cerveau au cours d'une période de temps aussi courte à l'échelle de l'évolution.*

Toutes ces questions sont fondamentales en paléontologie, en anthropologie et en archéologie, et demeurent largement sans réponse. En fait, il est difficile d'y répondre parce que les données dont on dispose sont fragmentaires ; les tissus mous ont disparu, laissant derrière eux surtout des ossements. Par conséquent, la structure et la fonction ne peuvent être reliées qu'indirectement. Mais une chose est claire : même si nous disposions de données de meilleure qualité, nous aurions quand

même besoin d'une théorie pour expliquer comment la morphologie apparaît et comment elle se modifie au cours de l'évolution.

Pourquoi ? Parce que la morphologie – la forme des cellules, des tissus, des organes et de l'animal tout entier – est à la base de tout comportement. Un grand nombre de résultats confirment cette conclusion, parfois de manière assez peu subtile et même triviale. Pour voler il faut avoir des ailes ; pour penser, il faut avoir un cerveau. Mais, à un autre niveau, la morphologie est extrêmement subtile. Ainsi, c'est une infime modification de la position du point d'insertion d'un muscle élévateur de la mâchoire des cichlidés qui a permis à ces poissons de déglutir sans lâcher leur appât. À partir de là, ceux de leurs descendants qui ont hérité de cette variante ont réussi à envahir diverses niches écologiques, et leur rayonnement adaptatif a été si explosif qu'ils ont devancé un grand nombre de concurrents. La morphologie est également importante pour nous autres humains : lorsque nous comparons les êtres humains et les chimpanzés au niveau génétique, nous constatons qu'ils sont identiques à 99 %. Mais les modifications morphologiques qui ont permis aux humains de devenir bipèdes, qui ont abouti à des modifications des points d'insertion des muscles de la mâchoire dans le crâne, à l'augmentation du volume crânien, à l'apparition d'une chambre supralaryngée contenant des organes vocaux, et à celle d'une partie du cortex cérébral appelée planum temporale, semblent avoir été essentielles à l'apparition de comportements spécifiquement humains.

Certains résultats incitent à penser qu'il existe une relation (non linéaire) entre la taille et la complexité du cerveau et celle des comportements. Par ailleurs, toute une masse de résultats indique qu'il existe un couplage entre les fonctions de régions spécifiques du cerveau et des capacités spécifiques. Et les données cliniques concernant les lésions cérébrales révèlent que des pertes de fonctions mentales identifiables et spécifiques surviennent lorsque des régions particulières du cerveau ont été lésées. Ces divers résultats suggèrent que, pour comprendre l'évolution de l'esprit et du comportement, nous devons d'abord avoir compris les bases de l'évolution morphologique.

De tels résultats tendent à confirmer l'hypothèse darwinienne selon laquelle les capacités mentales humaines sont issues de la sélection naturelle. Certaines données fossiles et archéologiques dénotent la présence d'activités mentales. Le fait que des humains aient enterré leurs morts, par exemple, peut être considéré comme

une preuve du fait qu'ils étaient conscients et même conscients d'eux-mêmes. Mais on peut également trouver d'autres arguments, probablement plus parlants, en analysant la structure du cerveau, la façon dont il fonctionne, l'origine de ses cellules, ce qu'elles ont en commun avec celles des autres espèces – et ce que le cerveau humain a de si différent et spécial.

On doit alors formuler une théorie de l'évolution et du développement de la morphologie animale et tissulaire. Il faut ensuite élaborer une théorie de la fonction cérébrale à partir de la première. Voilà la gigantesque tâche que l'on doit accomplir si l'on veut compléter le programme de Darwin. Ce qui rend cette entreprise fondamentalement intéressante, c'est qu'elle *n'est pas* spécifique du cerveau. Pour la mener à bien, nous devons montrer comment le développement (l'embryologie) s'articule avec l'évolution. Nous devons déterminer comment les gènes affectent la morphologie au cours du développement. Nous devons nous demander quelles sont les contraintes que ce processus impose à l'évolution – autrement dit, montrer que les règles du développement, qui sont elles-mêmes le résultat d'une évolution, ne peuvent être réalisées que d'un certain nombre de façons particulières.

Si ces connaissances sont indispensables, c'est parce que l'évolution est un processus historique, parce que, au cours du développement, seules certaines combinaisons d'événements sont susceptibles d'aboutir à des formes fonctionnelles, et parce que la forme du corps d'un animal est aussi essentielle au fonctionnement et à l'évolution de son cerveau que la forme et le fonctionnement du cerveau le sont au comportement de son corps.

Au cours du chapitre suivant, nous allons nous intéresser à la façon dont l'embryologie et l'évolution *interagissent* pour produire des cerveaux et des corps. Mais attention : ne vous attendez pas à des explications miraculeusement simples. Étant donné ce que nous savons sur l'évolution, l'existence d'une substance biochimique qui permette à elle seule de différencier les humains des singes est aussi peu probable que celle d'un gène de l'altruisme. Les liens entre morphologie et esprit, dont l'identification permettra d'achever le programme de Darwin, se révéleront plus indirects, plus détournés que cela. Mais leur imbrication les rendra d'autant plus fascinants. L'une des causes de cette fascination – et non des moindres – est d'ailleurs le fait qu'un œuf fécondé puisse donner naissance à un animal fonctionnel, cerveau compris. Cela suffira à justifier, je l'espère, le changement de rythme que

je m'apprête à opérer pour vous présenter un mini-cours de biologie moléculaire et d'embryologie modernes. Certaines des conclusions auxquelles nous parviendrons nous serons indispensables dans la partie suivante de ce livre, lorsque nous serons confrontés au problème de l'élaboration d'une théorie du cerveau compatible avec l'évolution et le développement.

Chapitre 6

La topobiologie : leçons d'embryologie

La poule et l'œuf enfin réunis.
Titre d'une critique de *Topobioloy*,
New York Times Review of Books,
22 janvier 1989.

Il peut paraître étrange de s'intéresser à l'embryologie dans un livre sur l'esprit. Les ovules et les spermatozoïdes ne semblent pas avoir d'esprit, pas plus que les embryons au tout début de la gestation. Mais, puisque nous savons que tout indique que les nouveaunés en ont un, aussi ténu soit-il, il semble raisonnable de se demander quelles sont les interactions qui permettent aux bases de la vie mentale de s'établir.

Pourquoi ce détour par des questions telles que la forme et la morphologie ? Pourquoi s'intéresser à des cellules, à des molécules, à l'ADN ? Tout simplement parce que les règles qui président à la construction de l'embryon conditionnent également la structure du cerveau. La formation de l'anatomie du cerveau dépend effectivement de l'action des muscles sur les os, de celle des nerfs sur la peau dans un certain ordre, et ainsi de suite – autrement dit, elle dépend du reste du phénotype. Comme je l'ai dit au chapitre précédent, si nous voulons déterminer à quel moment sont apparus divers aspects de l'esprit au cours de l'évolution, nous devons comprendre les liens existant entre développement et évolution.

Pour analyser le développement, j'aurai forcément besoin de faire appel à certains détails techniques. Je suggère aux lecteurs de lire une fois ce qui a trait à ces détails, de regarder les figures, puis de revenir au texte. Examinons tout d'abord quelques notions préliminaires (figure 6-1). Les cellules des organismes supérieurs (dits eucaryotes) ont un noyau qui contient de l'ADN – autrement dit, un patrimoine héréditaire. L'ADN se compose de longues chaînes de quatre types de molécules plus petites, appelées bases nucléotidiques, qui forment des séquences. Il n'existe que quatre types de bases : guanine, cytosine, adénine et thymine – G, C, A et T en abrégé. Par

exemple, la séquence d'un brin d'ADN pourrait s'écrire comme suit :... GTCGACCTGGCAGGTCAACGGATC... De plus, on sait aujourd'hui que chaque brin d'ADN se trouve enroulé avec un brin d'ADN dont la séquence est complémentaire de la sienne, c'est-à-dire, dans ce cas :... CAGCTGGACCGTCCAGTTGCCTAG... Remarquez que, dans cette « double hélice », chaque base s'apparie avec la base complémentaire : les G situées sur l'un des brins s'apparient avec les C situées sur l'autre, et les A avec les T. Par ailleurs, *à l'intérieur* de chaque brin, les G, les T, etc., sont liées les unes aux autres par des liaisons chimiques fortes, comme les perles d'un collier. En revanche, d'un brin à l'autre, les forces qui relient les G aux C et les A aux T sont des liaisons chimiques faibles, qui peuvent être rompues lors d'une augmentation de température, par exemple, ce qui permet aux brins de se séparer.

Les éléments clés sont les suivants : 1. En partant d'un brin d'ADN, on peut construire un second brin à partir des bases isolées, grâce à des protéines spéciales – des enzymes – qui catalysent, ou accélèrent, l'établissement de liaisons chimiques entre les bases du brin en cours de construction. L'ordre des bases dans le second brin est déterminé par l'ordre des bases dans le premier, puisque chaque base de ce brin s'apparie avec sa complémentaire sur le brin opposé. 2. Les séquences de trois bases (tout triplet contenant des G, C, A, T) sur un brin d'ADN représentent des mots de code, ou codons, qui commandent à la cellule de prendre les « briques » de construction des protéines – appelées acides aminés – correspondantes et de les utiliser pour former de longues chaînes, appelées polypeptides. Ensuite, en se repliant, ces chaînes polypeptidiques forment des protéines. Comme tous les mots de code ont trois « lettres », on peut construire 64 mots de code avec les quatre lettres correspondant aux quatre types de bases de l'ADN. Cependant, les protéines ne comportent que vingt acides aminés différents. Par conséquent, il est évident que certains mots de code ne codent pas pour des acides aminés, et que certains autres sont tout simplement redondants. Tout fragment d'ADN dont la longueur et la séquence lui permettent de spécifier une protéine donnée est un gène. 3. Lorsqu'une cellule se divise, elle recopie l'ADN contenu dans l'un des brins de façon à fournir une nouvelle molécule d'ADN à chacune de ses deux cellules-filles. Normalement, chaque copie comportera exactement la même séquence de mots de code. Cependant, en cas d'erreur, ou lorsqu'un des brins d'ADN est sectionné – disons, par exemple, par le rayonnement cosmique –, il arrive que les enzymes de réplication et de

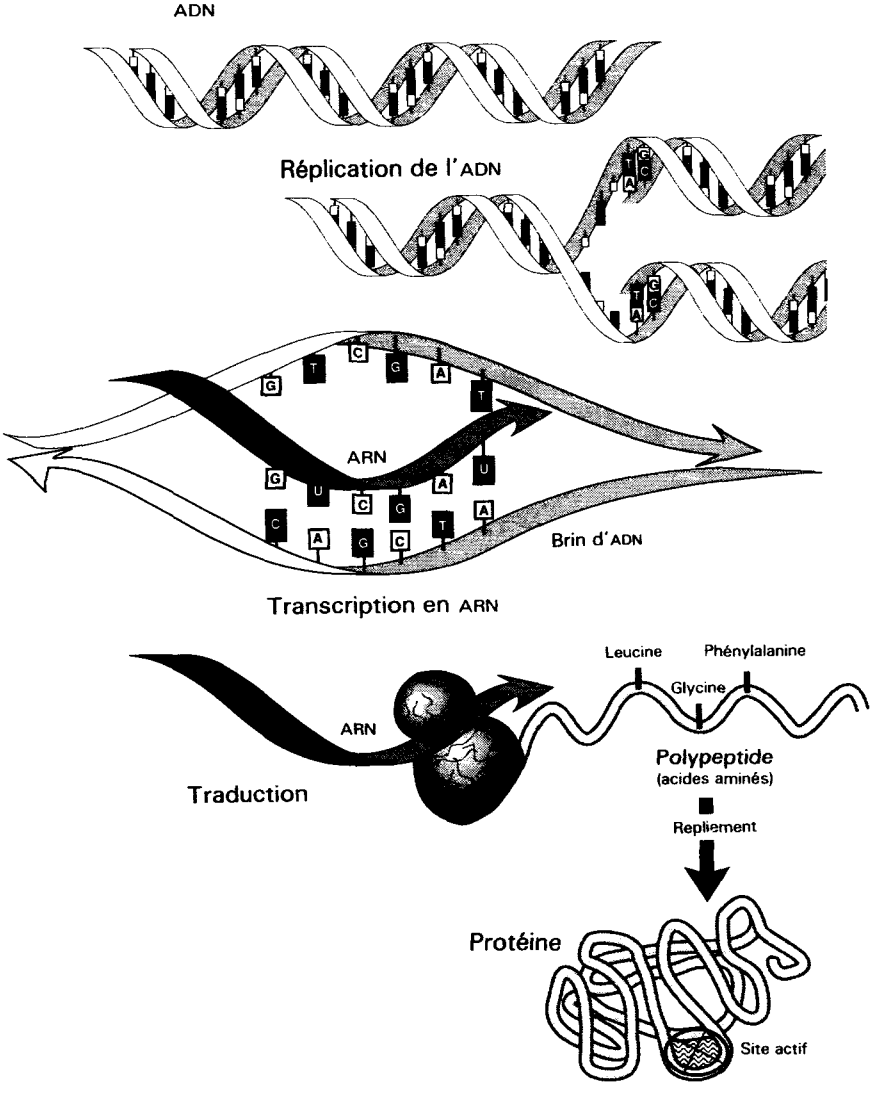


FIGURE 6-1 : *Du décodage de l'information génétique à la fabrication des protéines (un mini-cours de biologie moléculaire). L'ADN comporte deux brins joints par des forces faibles s'exerçant entre bases nucléotidiques complémentaires. La guanine (G) s'apparie avec la cytosine (C), tandis que l'adénine (A) s'apparie avec la thymine (T). La séquence de bases contenue dans l'un des brins – liée par des forces beaucoup plus intenses – peut être lue comme une suite de mots de code de trois lettres chacun, spécifiant chacun un acide aminé particulier (les acides aminés sont les « briques » de construction des protéines). L'ADN lui-même demeure enfermé dans le noyau des cellules, mais l'information qu'il contient est transportée vers un autre site sous la forme d'un brin d'ARN. Celui-ci est construit par des enzymes spéciales, chargées de transcrire la séquence d'ADN. (Le code de l'ARN est le même, à ceci près qu'il fait appel à une base différente, l'uracil (U), qui prend la place de la thymine ; voici quelques mots code typiques : UUU = phénylalanine, CUU = leucine, GGC = glycine, etc.) Cette séquence codée d'ARN est ensuite « traduite » sur une structure spécialisée appelée ribosome, où d'autres enzymes relient chaque acide aminé tour à tour, à mesure que leurs codes sont lus sur l'ARN. (Les acides aminés sont transportés sur ce site par un autre type d'ARN.) Le polypeptide résultant de l'enchaînement des acides aminés se replie pour adopter une forme tridimensionnelle qui dépend de sa séquence. Une seule modification, ou mutation, dans la séquence initiale d'ADN peut suffire à modifier la séquence d'acides aminés, ce qui peut à son tour entraîner une modification de la forme de la protéine. Cela peut modifier sa fonction, qui est assurée par un site actif de sa structure repliée. Lorsqu'une cellule se divise, les deux brins d'ADN se séparent et une enzyme recopie chacun des deux brins afin de donner le même ADN à chacune des deux cellules-filles. Ce processus est appelé réplication.*

réparation de l'ADN ne recopient pas fidèlement le brin de départ. Voilà l'une des façons par lesquelles une mutation est susceptible de survenir dans un gène et de modifier ses mots de code.

Ce sont Oswald Avery et ses collègues de l'Institut Rockefeller (États-Unis) qui ont découvert que l'ADN constituait le matériel génétique de la cellule. Ces chercheurs ont annoncé leur résultat dans un article extraordinairement important, qui a été publié dans le *Journal of Experimental Medicine* en 1944. Curieusement, l'article n'a suscité ni l'adhésion immédiate ni l'intérêt général. À l'époque, je jouais de la musique avec Stuart Elliot, un microbiologiste qui travaillait avec Avery et aussi avec Fred Griffith, qui, en Angleterre était l'un des pionniers en ce domaine. Un jour de 1964, Stuart me suggéra de parler à Peyton Rous, l'éditeur du *Journal of Experimental Medicine,* pour lui proposer de publier à nouveau l'article fondamental d'Avery et de ses collègues. Cela permettrait, selon Stuart, de commémorer convenablement à la fois le vingt-cinquième anniversaire de la mort de Griffith et le vingtième anniversaire de la publication de l'article d'Avery.

Rous (qui, à plus de 80 ans, reçut le Prix Nobel pour les travaux qu'il avait effectués au début de sa carrière sur les virus à l'origine de cancers) promit de « soumettre l'idée au conseil éditorial du journal ». Je n'avais plus eu de nouvelles de lui depuis six semaines lorsque, par hasard, je le rencontrai qui descendait d'un autobus. Je lui demandai ce qu'il en était. Il s'arrêta, me regarda d'un air grave, et me dit : « Ah, oui. J'en ai parlé au conseil, et ils ont trouvé l'idée extrêmement vulgaire. » Surpris, je parcourus cent mètres en silence à ses côtés, et je l'entendis alors dire doucement : « Je n'ai jamais beaucoup aimé Avery. » Lorsque je lui demandai pourquoi, il me répondit : « Que penseriez-vous d'un homme qui a reçu une médaille de la Royal Society et qui n'est jamais allé la chercher ? »

Avery ne reçut pas le Prix Nobel. L'importance de son travail ne fut vraiment reconnue que quelques années après sa mort. Encore aujourd'hui, malgré notre compréhension de la biologie moléculaire de l'ADN, nous avons du mal à imaginer à quel point sa découverte fut capitale. En science comme ailleurs, les mérites ne sont pas toujours justement reconnus.

Ce petit précis de biologie moléculaire nous permet d'aborder la relation entre génotype (l'ensemble des gènes d'un organisme) et phénotype. Les gènes sont de longues chaînes de mots de code

comportant des signaux de départ et d'arrêt (qui font également partie du code). Par l'intermédiaire de la machinerie cellulaire, l'ADN est transcrit en une autre longue chaîne de bases quelque peu différentes, appelée ARN. L'ARN est ensuite transporté hors du noyau de la cellule afin d'être lu par un dispositif cellulaire qui rappelle la tête de lecture d'un magnétophone. Celui-ci assemble les acides aminés correspondant à la séquence codante, dans l'ordre prescrit, et construit un polypeptide parfois long de quelques centaines d'acides aminés (figure 6-1).

Une fois sa construction achevée, ce polypeptide se replie sur lui-même et adopte une forme complexe, devenant ainsi une protéine plus ou moins compacte (figure 6-2). L'ordre dans lequel les acides aminés sont rangés sur la chaîne détermine la forme de la protéine. C'est cette *forme* qui confère à la protéine ses propriétés phénotypiques et fonctionnelles ; par exemple, certaines formes permettront à des protéines de mieux s'encastrer (à la manière des blocs d'un jeu de construction) dans d'autres protéines pour former des structures cellulaires, tandis que d'autres formes leur permettront plutôt de se lier à des composés chimiques pour changer la vitesse à laquelle ces composés réagissent. Comme je l'ai déjà évoqué, cette dernière caractéristique constitue la principale propriété des enzymes. Résumons :

1. L'ADN « fabrique » de l'ARN, qui à son tour « fabrique » des protéines (le mot entre guillemets veut dire « spécifie », car c'est en fait la cellule qui fabrique ces substances chimiques).

2. La forme de chaque protéine dépend de sa séquence d'acides aminés, qui dépend à son tour de la séquence initiale de mots de code contenue dans l'ADN correspondant.

3. La fonction d'une protéine dépend de sa forme.

Étant donné qu'une grande partie du phénotype d'un organisme dépend des propriétés de ses protéines, ces règles pourraient sembler rendre compte non seulement de la forme des protéines, mais aussi, par extension, de la forme des animaux.

Hélas, les choses ne sont pas si simples, car ce n'est pas un assemblage de *protéines*, mais un assemblage de *cellules* qui donne un embryon. La forme de l'embryon et celle de ses tissus – et notamment celle de son cerveau – proviennent de la forme des divers ensembles de différents types de cellules, chaque type ayant des protéines différentes (les différences sont dues au fait que différentes combinaisons de gènes s'expriment dans chaque type de cellules).

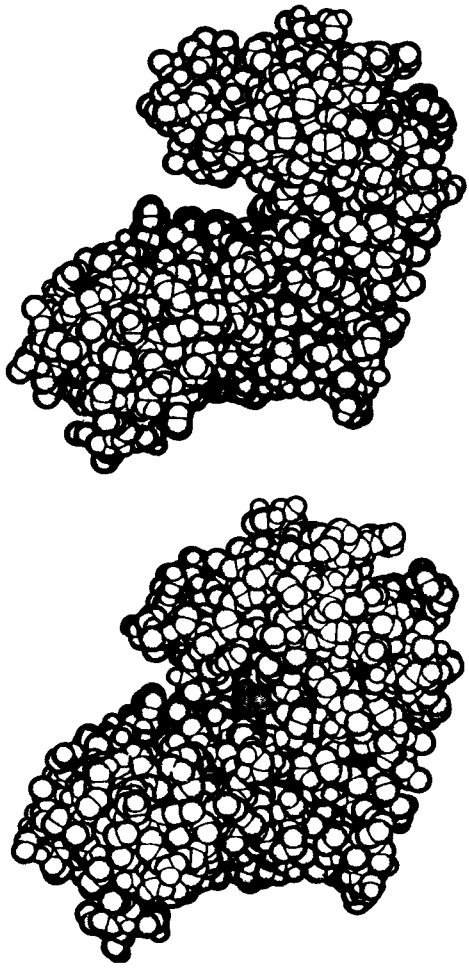


FIGURE 6-2

Repliement et fonction des protéines : un exemple. Cette protéine repliée, appelée hexokinase, *est capable de se lier au sucre glucose au niveau de la cavité qui constitue le site actif de sa fonction enzymatique. Cette fonction consiste à catalyser une réaction chimique métabolique au cours de laquelle des dérivés phosphorés se lient au glucose.* En haut : *la protéine en l'absence de glucose.* En bas : *lorsqu'on ajoute du glucose, la cavité se referme autour de lui, le retenant fermement.*

Par conséquent, nous devons à présent nous demander comment les cellules s'y prennent. Mais ce faisant, nous ne devons pas perdre de vue un élément clé – le fait qu'en dernière analyse, la forme d'un animal dépend effectivement de l'ordre dans lequel sont rangés les mots de code dans son ADN. De plus, les changements de forme au cours de l'évolution doivent être dus à des mutations qui ont

modifié l'ordre des mots de code dans l'ADN d'un ancêtre. Par conséquent, les questions auxquelles nous devons répondre sont :

1. Comment un code génétique unidimensionnel réussit-il à spécifier, non pas seulement la forme d'une molécule protéique tridimensionnelle, mais aussi celle d'un animal tridimensionnel tout entier ?

2. Comment expliquer que des modifications survenant au cours du temps dans les processus de développement qui conduisent à ces formes soient à l'origine de l'apparition de nouvelles formes ?

Pour répondre à ces questions de manière provisoire, j'ai écrit un livre intitulé *Topobiology*, dans lequel j'explique, entre autres, comment le cerveau a pu apparaître au cours de l'évolution. *Topos* signifie lieu, et le titre fait allusion au fait qu'un grand nombre des transactions cellulaires qui déterminent la forme d'un organisme dépendent de la position des cellules dans l'embryon : elles n'ont lieu que lorsqu'une cellule se trouve entourée d'autres cellules à un endroit particulier. Examinons à présent certains de ces phénomènes dépendant de la position qui aboutissent à la formation d'un embryon et de ses organes – et en particulier à celle de son cerveau.

L'embryon se forme lorsqu'un spermatozoïde, contenant de l'ADN issu d'un mâle, féconde un ovule portant l'ADN issu d'une femelle. (À ce propos, les cellules germinales – les spermatozoïdes et les ovules – sont extrêmement diversifiées, chacune étant susceptible de contenir des gènes porteurs de différentes mutations.) Il existe beaucoup de gènes, et chacun d'entre eux possède une longue séquence de mots de code. La fusion du spermatozoïde et de l'ovule donne lieu à ce qu'on appelle un zygote. Le zygote possède désormais des gènes provenant de ses deux parents, et il commence à subir une série de divisions, ou clivages, qui aboutissent à 2, 4, 8,..., 2^n cellules. La masse de cellules-filles qui en émerge a habituellement la forme d'une boule (chez les oiseaux, cependant, elle peut être plane).

Avant de poursuivre ma description, je dois m'arrêter un instant pour donner quelques précisions sur l'activité des cellules :

1. Les cellules *se divisent*, et ce faisant, transmettent la même quantité et le même type d'ADN à leurs cellules-filles.

2. Les cellules *migrent*, en se détachant par couches ou feuillets appelées épithéliums, et forment un ensemble, mouvant et peu serré, appelé mésenchyme. (Les couches peuvent également se déplacer

en s'enroulant sur elles-mêmes pour former des tubes sans que les contacts entre leurs cellules se défassent.)

3. Les cellules *meurent* à des endroits précis.

4. Les cellules *adhèrent* les unes aux autres, comme nous l'avons déjà vu, ou bien elles perdent leur adhérence et migrent vers d'autres positions. Cette migration se déroule soit à la surface d'autres cellules pour aboutir à la formation de couches, soit sur une matrice de molécules libérées par les cellules. Ensuite, les cellules adhèrent à nouveau entre elles, et forment de nouvelles combinaisons.

5. Les cellules se *différencient* ; autrement dit, elles expriment différentes combinaisons des gènes présents dans leur noyau. Elles peuvent le faire à n'importe quel moment et à n'importe quel endroit, mais seulement lorsqu'elles ont reçu les signaux appropriés. Les signaux appropriés ne sont émis qu'à certains endroits de l'embryon en formation. Le processus d'expression différentielle des gènes est appelé différenciation. C'est lui qui rend les cellules du foie différentes de celles de la peau, celles de la peau différentes de celles du cerveau, et ainsi de suite. Différenciation veut donc dire production d'ensembles spécifiques de protéines, puisque les gènes codant pour certaines protéines sont activés tandis que d'autres restent inactifs. Chaque cellule d'un type donné possède de nombreuses protéines, mais elle n'en a que quelques-unes en commun avec les autres types de cellules.

Nous pouvons à présent reprendre notre description de la manière dont se construit l'embryon. Considérons l'embryon de poulet (figure 6-3). Les divisions cellulaires successives aboutissent à la formation d'une plaque de cellules appelée blastoderme, qui contient plus de 100 000 cellules. À ce stade, les cellules situées de part et d'autre de la ligne médiane, dans la partie postérieure du blastoderme, se détachent et migrent à travers la portion centrale de celui-ci, appelée gouttière primitive. Elles finissent ainsi par se retrouver sous le blastoderme, où elles adhèrent les unes aux autres pour former une couche appelée mésoderme. Trois couches distinctes – l'ectoderme, le mésoderme et l'endoderme – vont se former à l'issue de ce processus, appelé gastrulation.

À ce stade, il se produit un événement étonnant, qui fait intervenir à la fois la position et la communication cellulaires. Il s'agit de ce qu'on appelle l'*induction embryonnaire*, qui résulte du passage de signaux d'un ensemble de cellules situées dans l'une des couches vers un ensemble de cellules situées dans une autre couche. Les

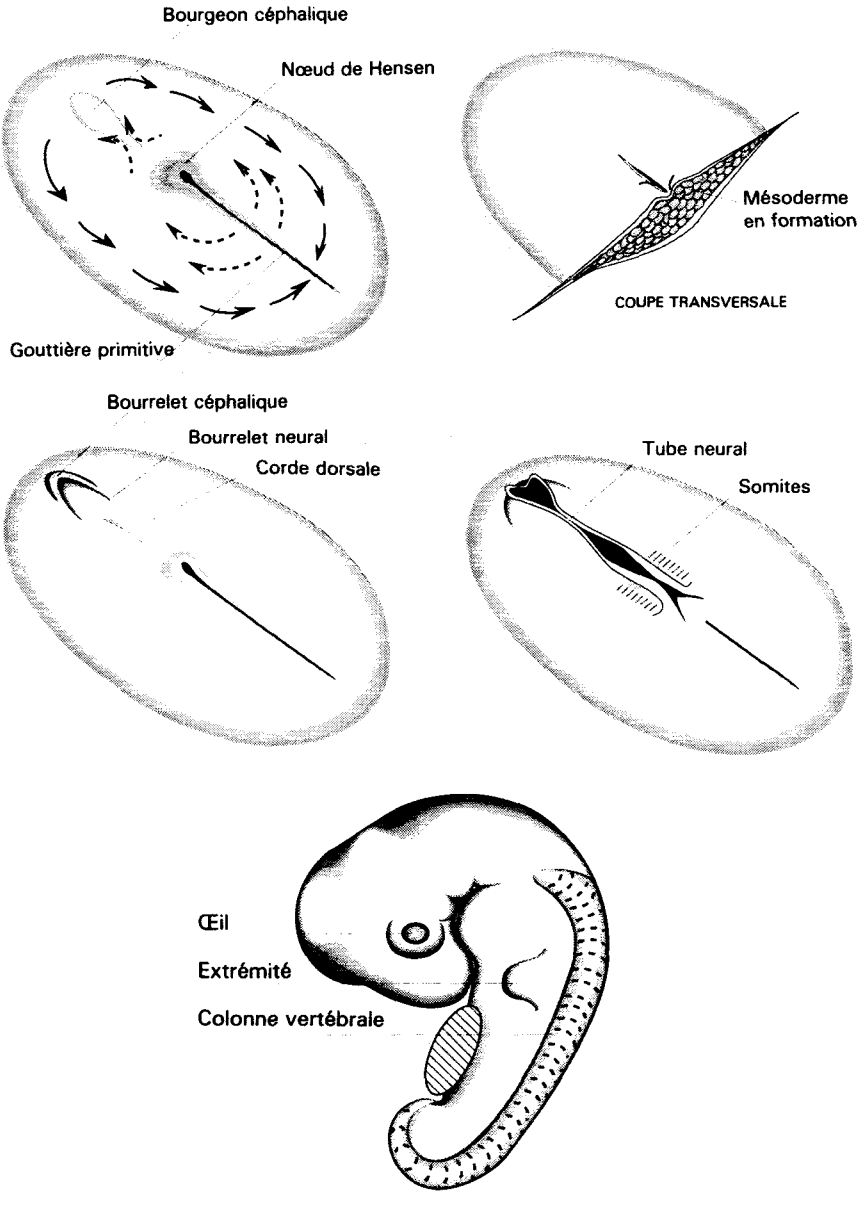


FIGURE 6-3 : *Le début du développement de l'embryon de poulet à partir d'une plaque de cellules aboutit à la formation de plusieurs couches au cours d'un processus appelé gastrulation. Les cellules traversent la gouttière primitive* (en haut à gauche)) *et forment des couches* (en haut à droite). *L'axe central de la plaque se replie ensuite pour former le tube neural, qui donnera ultérieurement lieu au système nerveux. Peu après, des cellules des couches inférieures se séparent* (au centre à gauche) *et forment des structures segmentées appelées somites* (au centre à droite). *Le système nerveux continue à se développer à mesure que les cellules du tube neural font croître leurs prolongements* (illustrés sur la figure 3-3). *Il en résulte un embryon qui commence à ressembler à un animal* (en bas).

cellules du mésoderme envoient des signaux à celles de l'ectoderme, ce qui induit une différenciation dépendant de la position (c'est-à-dire topobiologique) d'une section centrale de cellules. Ces cellules vont alors former la plaque neurale. Les cellules situées hors des frontières de la plaque neurale donneront naissance à la peau ; celles de la plaque proprement dite formeront le tube neural (figure 6-3) et par la suite, le système nerveux. Ce dernier sera issu des diverses couches qui se seront formées, en s'enroulant sur elles-mêmes, à l'intérieur du tube. Cela définit non seulement l'axe de l'embryon, mais aussi la position de la tête de l'animal et les signaux dépendant de la position qui serviront aux processus d'induction futurs.

Il est à remarquer que ces processus primaires de division, de migration, de mort, d'adhérence et d'induction cellulaires varient en fonction du temps et de la position. De fait, la clé de la consolidation de ces événements réside dans leur *coordination*, car c'est elle qui permet à de nouveaux signaux inductifs d'être émis, qui, à leur tour, conduiront à encore d'autres modifications à un endroit donné, lequel est le résultat de modifications passées. À l'issue de certaines inductions secondaires, par exemple, les cellules nerveuses du tube neural projetteront des fibres qui formeront des réseaux, chacun étant spécifique d'une région du cerveau ou de la moelle épinière (voir la figure 3-3). D'autres cellules formeront les yeux, l'intestin ou les reins. Et tout se déroulera de manière à produire la forme caractéristique de l'espèce – dans ce cas, un poulet.

Cette question de la forme est essentielle. En effet, elle revient à dire que des *combinaisons* de gènes doivent agir pour donner naissance à une forme héréditaire, caractéristique de l'espèce. Elle revient également à dire que les phénomènes mécaniques qui aboutissent à la réorganisation et à la spécialisation des cellules doivent être coordonnés avec l'expression séquentielle des gènes. *Il s'agit là de la condition nécessaire de la topobiologie.* Elle permet de comprendre pourquoi le fait que les gènes spécifient la forme des protéines ne suffit pas ; en fait, les véritables moteurs du développement, ce sont les *cellules* elles-mêmes, qui se déplacent et qui meurent de façon imprévisible. La fabrication de protéines ou de surfaces cellulaires qui s'encastrent les unes dans les autres à la manière des pièces d'un jeu de Lego – chacune étant spécifique d'une cellule donnée – ne suffit pas pour expliquer comment les gènes spécifient la forme d'un organisme. En effet, bien que toutes les cellules des embryons d'une espèce se ressemblent *en moyenne*,

les déplacements et la mort d'une cellule *donnée* à un endroit *donné* sont d'ordre statistique, et la position effective de la cellule ne peut donc pas être spécifiée d'avance par le code contenu dans un gène.

Mais alors, *qu'est-ce* qui explique que cette merveilleuse séquence de ballets cellulaires et de signaux aboutisse à une forme ? L'un des indices révélateurs se trouve au niveau moléculaire, au niveau de molécules dites morphorégulatrices, lesquelles régissent l'adhérence et le mouvement (figure 6-4). Ces protéines sont spécifiées par certains ensembles de gènes à des endroits spécifiques de l'embryon. Leur principale fonction consiste à provoquer l'adhérence des cellules à un substrat ou à relier des cellules entre elles de telle sorte qu'elles forment des couches appelées épithéliums. Ces molécules appartiennent à l'une des trois familles suivantes : les molécules d'adhérence cellulaire (CAM), qui relient directement les cellules entre elles ; les molécules d'adhérence au substrat (SAM), qui relient les cellules de façon indirecte, en leur fournissant une matrice, ou base, sur laquelle se déplacer ; et les molécules de jonction cellulaire (CJM), qui permettent aux cellules reliées par des CAM de former des couches épithéliales.

L'argument clé est donc le suivant : l'activation de gènes correspondant à des sous-ensembles de molécules morphorégulatrices *modifie* la mécanique des cellules et des épithéliums. Ce processus dépend de la chimie de chaque cellule, qui agit sur les structures internes affectant la forme et le mouvement cellulaires et qui porte donc le nom de mécano-chimie. Une combinaison spécifique de CAM et de SAM, par exemple, permettra à certaines cellules de se déplacer et régira les phénomènes mécano-chimiques qui font que les couches formées par adhérence des cellules entre elles se replient. Elle empêchera même le déplacement de certaines cellules à certains endroits. Du fait de cette action tour à tour permissive et restrictive, l'expression des CAM (qui sont liées à la surface des cellules) et celle des SAM (qui sont déposées par des cellules afin d'interagir avec les surfaces d'autres cellules en des endroits particuliers) peuvent modifier les combinaisons de cellules présentes à un endroit donné, et donc conduire à des formes différentes. Comme ce sont des gènes particuliers qui commandent la formation de CAM et de SAM particuliers, cette action peut être héréditaire au sein d'une espèce donnée, et donc déterminer son mécanisme morphogénique.

Cependant, cela ne suffit pas encore. D'autres événements doivent se produire parallèlement à l'expression des CAM, des SAM et des

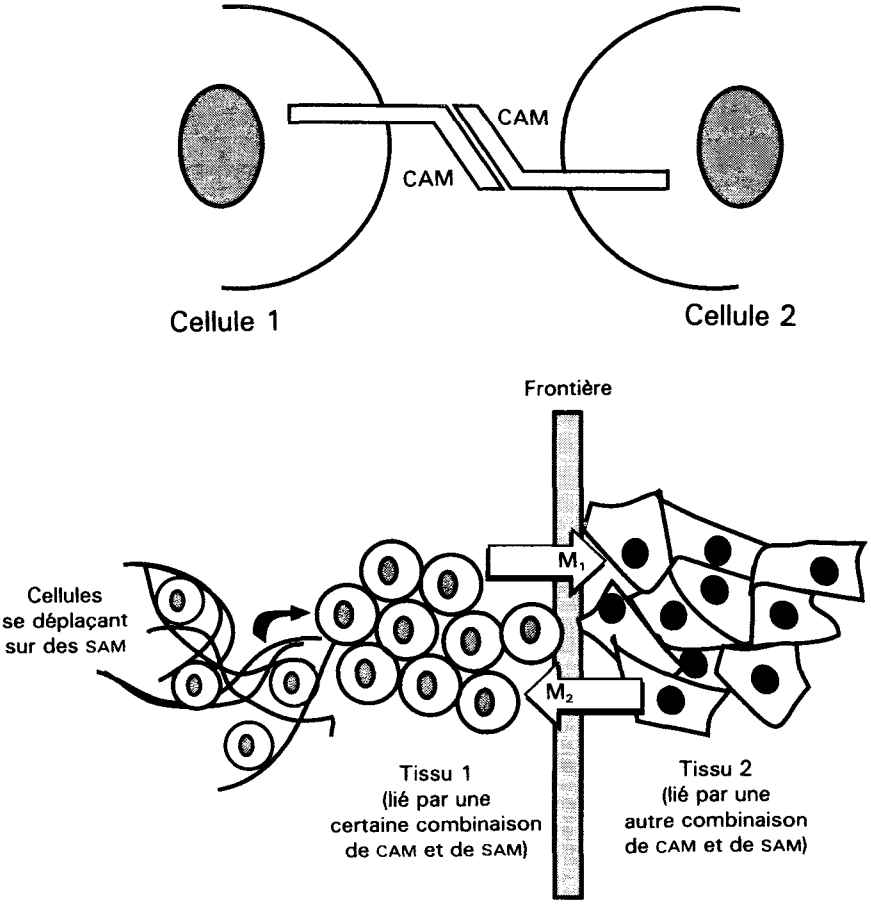


FIGURE 6-4

L'adhérence cellulaire. En haut : *les cellules se lient entre elles au moyen de protéines particulières présentes dans leur membrane extérieure, appelées molécules d'adhérence cellulaire (CAM pour* cellular adhesion molecules*).* En bas : *d'autres molécules, appelées molécules d'adhérence au substrat (SAM pour* substrate adhesion molecules*) forment des matrices extracellulaires sur lesquelles les cellules se déplacent et s'appuient. Les CAM et les SAM règlent la façon dont les cellules s'assemblent et se séparent, et autorisent ou interdisent les déplacements cellulaires. Les processus sous-jacents à l'apparition de la forme, comme on le voit sur la figure précédente, sont sous le contrôle de gènes spécifiques qui s'expriment là où des signaux moléculaires (morphogènes, M_1 et M_2) sont échangés entre des ensembles adjacents de cellules. Le processus topobiologique est complexe, mais l'idée de base est simple : les cellules se déplacent ou se fixent à un endroit donné ; suite à l'activation ou à l'inactivation d'un ensemble de gènes, les cellules sont soit libérées, soit retenues, et elles émettent alors de nouveaux signaux qui mènent à de nouvelles combinaisons. Cela entraîne des modifications supplémentaires de l'expression des CAM et des SAM, et la modification des types cellulaires et tissulaires par des morphogènes. Ces processus donnent naissance aux formes et aux types tissulaires qui constituent l'embryon.*

CJM de types donnés, en quantités données, à des endroits donnés. Une fois achevée la construction d'un site sous l'effet des divers signaux inductifs, de nouvelles combinaisons de signaux doivent dire aux cellules qui se trouvent là quelles sont les nouvelles CAM à activer et les anciennes à inactiver. Ces signaux consistent en de petites molécules ou facteurs de croissance qui interagissent directement ou indirectement avec les gènes correspondants afin de contrôler leur expression (figure 6-4).

De même, il faut que quelque chose induise ou empêche la différenciation des cellules aux endroits appropriés. Les biologistes moléculaires ont identifié des gènes d'un type particulier, appelés gènes homéotiques. Les gènes homéotiques spécifient des protéines qui, en se liant à des portions d'autres gènes, règlent la production des protéines spécifiées par ces gènes. Ainsi, les gènes homéotiques contrôlent les phénomènes de différenciation qui aboutissent à la formation de parties du corps telles que les ailes, les yeux, ou des portions de la colonne vertébrale. Chez la mouche du vinaigre, ou drosophile, par exemple, une mutation survenant dans un gène homéotique peut provoquer la croissance d'une patte à la place d'une antenne (figure 6-5). Les gènes homéotiques s'expriment en gradients à travers le corps de l'animal, en général de l'avant vers l'arrière, et dans des régions particulières.

Résumons encore : les cellules expriment, dans le temps et dans l'espace, des gènes qui régissent la production de molécules morphorégulatrices, qui à leur tour contrôlent les mouvements cellulaires et l'adhérence entre cellules. Ces actions placent des groupes de cellules à proximité les uns des autres et leur permettent d'échanger encore d'autres signaux inductifs. Ces signaux modifient l'expression des gènes homéotiques, qui ensuite vont modifier l'expression d'autres gènes. Les acteurs clés de cette cascade topobiologique sont les cellules, qui se déplacent, meurent, se divisent, libèrent des signaux inductifs ou morphogènes, se lient entre elles pour former de nouvelles couches et reprennent ensuite ce processus sous des formes plus ou moins semblables. Les gènes contrôlent l'ensemble indirectement en sélectionnant les produits morphorégulateurs ou homéotiques qui seront exprimés. Mais le destin microscopique concret de chaque cellule de l'embryon dépend d'événements épigénétiques, qui à leur tour dépendent de l'histoire particulière de cette cellule au cours du développement.

La morphologie est le produit final de cet ensemble d'événements.

Et puisque chaque espèce possède une combinaison particulière de gènes, dont la fréquence est établie en moyenne par sélection naturelle au cours de l'évolution, chaque espèce possède une forme ou structure tissulaire plus ou moins caractéristique.

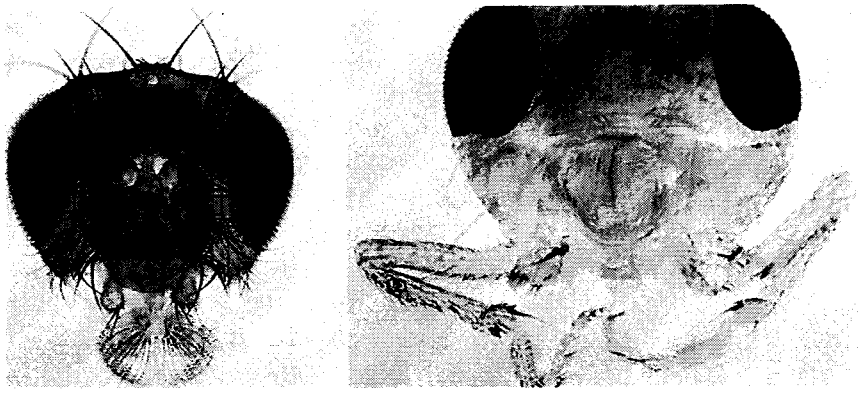


FIGURE 6-5

Un exemple de topobiologie aberrante. Les gènes homéotiques, qui règlent l'activité d'autres gènes, s'expriment en des sites précis au cours du développement embryonnaire. Chez la drosophile, ou mouche du vinaigre, lorsqu'un gène homéotique subit une mutation, une partie du corps – par exemple une patte – peut se substituer à une antenne. On voit, sur la photographie de gauche, la tête d'une mouche normale ; sur celle de droite, on voit une mouche mutante, portant des pattes à la place des antennes.

Cette description fournit une réponse provisoire à la question de savoir comment un code génétique unidimensionnel est capable de spécifier un animal tridimensionnel. Elle permet également d'imaginer comment l'évolution conduit à des modifications morphologiques relativement grandes et rapides. Supposons, par exemple, qu'au cours de l'évolution une mutation ait empêché une molécule morphorégulatrice d'apparaître au moment habituel ou de se lier de façon habituelle, retardant son expression en quantités suffisantes jusqu'à ce que les cellules de la région embryonnaire concernée se soient divisées plus que de coutume. Il pourrait en résulter une structure plus grande, de forme différente. Et si les nouveaux animaux, de taille et de forme différentes, se révélaient mieux adaptés à un environnement donné, la sélection naturelle favoriserait la reproduction différentielle de ces animaux. Cela aboutirait à une augmentation de la fréquence du gène mutant dans cette population, et davantage d'animaux naîtraient avec les nouvelles caractéristiques de taille et de forme.

Pourquoi m'être autant attardé sur cette question ? Pour deux raisons : parce que le système nerveux et le cerveau sont formés par des processus du type de ceux que je viens de décrire, et parce que l'émission de signaux dans le système nerveux est topobiologique (voir le chapitre 3). Les cartes du système nerveux, qui résultent du fait que les cellules nerveuses projettent leurs prolongements vers d'autres régions cellulaires au cours du développement, sont parmi les structures topobiologiques les plus remarquables qui soient. On sait aujourd'hui que leur formation dépend de molécules morphorégulatrices. De plus, les cartes dépendent souvent de la mort sélective des cellules qui entrent en compétition pour les fabriquer. Et elles requièrent aussi l'existence de processus de signalisation particuliers, capables de localiser les ramifications des cellules nerveuses voisines actives afin qu'elles se retrouvent également voisines dans la lointaine carte qui constitue leur cible. Les faits suggèrent que, lorsque l'activité chimique ou électrique des cellules nerveuses est bloquée, leurs ramifications ne forment pas des cartes bien ordonnées dans des régions éloignées (voir la figure 3-4).

Imaginons à présent un scénario épigénétique au cours duquel des couches de cellules nerveuses dans le cerveau en formation constituent un voisinage. Dans ce voisinage, les cellules échangent des signaux à mesure qu'elles établissent des liens par l'intermédiaire des CAM et des CJM. Elles projettent une profusion de prolongements, parfois structurés en paquets appelés faisceaux. Lorsque ces prolongements atteignent d'autres voisinages et d'autres couches, ils stimulent des cellules cibles. Celles-ci, à leur tour, libèrent des substances ou des signaux diffusibles qui, s'ils se trouvent être corrélés avec ceux des prolongements entrants, permettent à ces prolongements de se ramifier et de former des connexions avec les cellules cibles. Quant aux prolongements entrants dont les signaux ne sont pas corrélés avec les cibles, soit ils passent sans s'arrêter, soit ils se rétractent. D'ailleurs, lorsqu'un prolongement ne parvient pas à localiser sa cible, il arrive même que la cellule qui l'a émis meure. Finalement, à mesure que la croissance et la sélection s'opèrent, une carte neuronale ayant une fonction peut se former. Le nombre de cellules qui naissent, meurent et s'intègrent est énorme. Et la situation d'ensemble est totalement dynamique : elle dépend de nombreux signaux, des gènes, des protéines, et aussi des déplacements, des divisions et de la mort des cellules, et tous ces éléments interagissent à de multiples niveaux.

Remarquez les principales caractéristiques de ce scénario. Il est topobiologique, c'est-à-dire qu'il est dépendant de la position. Pour que des événements surviennent à un endroit, il faut que d'autres événements aient eu lieu au préalable à d'autres endroits. Mais par ailleurs, il s'agit d'un scénario intrinsèquement dynamique, plastique, variable au niveau des unités fondamentales – les cellules. Même chez des jumeaux génétiquement identiques, on ne trouve pas exactement la même structure de cellules nerveuses au même endroit et au même moment. Néanmoins, l'ensemble est quand même spécifique de l'espèce, parce que les contraintes *globales* agissant sur les gènes sont caractéristiques de cette espèce.

Les événements que j'ai décrits ici sont des événements sélectifs. Certaines structures cellulaires sont topobiologiquement sélectionnées au sein d'une masse cellulaire contenant diverses variantes. C'est flagrant dans le cas du système nerveux. La sélection garantit non seulement la formation d'une structure commune à toute l'espèce, mais aboutit également à une diversité individuelle au niveau des plus fins réseaux neuronaux. J'ai déjà évoqué le fait que la diversité ou variabilité des connexions à un endroit donné du système nerveux va à l'encontre de l'idée selon laquelle le cerveau fonctionnerait comme un ordinateur. En effet, la diversité résulte inévitablement de la nature *dynamique* des phénomènes topobiologiques. L'existence d'une diversité au niveau de chaque animal est fondamentale. En fait, il s'agit peut-être là de l'une des principales caractéristiques de la morphologie qui a donné lieu à l'esprit.

Mais nous n'y sommes pas encore ! D'abord, nous devons nous demander comment les systèmes biologiques font pour exécuter des tâches de reconnaissance – comment est-ce qu'ils parviennent, malgré l'absence de tout transfert de messages préexistants et spécifiquement codés, à distinguer spécifiquement les choses.

Chapitre 7

Les problèmes revus et corrigés

> *L'homme est à lui-même le plus prodigieux objet de la nature ; car il ne peut concevoir ce que c'est que corps, et encore moins ce que c'est qu'esprit, et moins qu'aucune chose comme un corps peut être uni avec un esprit. C'est là le comble de ses difficultés, et cependant c'est son propre être.*
>
> Blaise Pascal

Le fait de replacer l'esprit dans la nature a déclenché une série de crises scientifiques, car les données sur le cerveau, l'esprit et le comportement ne correspondent pas aux images que nous avons l'habitude d'utiliser pour les décrire. Je sais que bien des gens trouvent cette affirmation trop audacieuse, injustifiée, prématurée (« Avec davantage de données, on y verra plus clair ! »), ou tout simplement malsaine. Je pense au contraire que le meilleur moment pour travailler dans un domaine scientifique donné, c'est justement lorsqu'il est en crise. C'est à ces moments-là que l'on se voit obligé de trouver une nouvelle manière de regarder les données, ou une nouvelle théorie, ou encore une nouvelle technique, pour résoudre un paradoxe apparent. L'une des crises les plus frappantes de la science moderne, par exemple, a éclaté lorsque l'on a compris que l'application des lois classiques de la physique au rayonnement émis par un bloc de métal chauffé contenant une cavité (un « corps noir ») conduisait à une situation inacceptable aux courtes longueurs d'onde et aux hautes énergies ; en effet, dans cette situation dite de « catastrophe ultraviolette », l'énergie rayonnée devient infinie. Il a fallu résoudre le problème. Et c'est Max Planck qui a fourni la solution en suggérant que le rayonnement d'énergie issu d'un corps chauffé n'est pas continu, mais plutôt distribué en paquets, ou quanta.

Cependant, les crises suscitées par la question de l'esprit ne sont pas aussi claires. Cela n'a rien d'étonnant vu la subtilité et les multiples niveaux de développement du cerveau, de l'action du cerveau et des activités mentales. Cette affaire commence par les molécules et passe ensuite aux gènes. Elle fait intervenir d'énormes quantités de cellules présentant une activité électrique et une diver-

sité chimique, une anatomie extrêmement complexe – faite de *blobs* [1] et de couches hautement connectés – et des cartes qui reçoivent des signaux provenant des entrées sensorielles et en envoient aux sorties motrices. Ces structures sont continuellement soumises à des modifications électriques et chimiques, engendrées par les mouvements de l'animal, et en suscitent d'autres à leur tour. Ces mouvements sont eux-mêmes conditionnés par la forme et la structure de l'animal, et ils aboutissent à des comportements. Certains de ces comportements font intervenir des communications avec la mémoire de l'animal, qui à son tour est affectée par ses propres produits.

Tout cela découle de l'évolution – c'est-à-dire de l'action de la sélection naturelle durant des centaines de millions d'années. Il n'est donc pas surprenant que les crises de la neurobiologie et de la psychologie ne soient pas aussi précises ni aussi évidentes que celles de la physique. La complexité même des éléments en jeu est beaucoup plus grande qu'en physique. Pourtant, comme je l'ai montré chapitre 3, les crises deviennent claires dès lors que l'on est prêt à prendre en compte les différents niveaux en jeu pour tenter de relier la structure à la fonction.

Comment réconcilier les interactions des divers niveaux et résoudre les crises de structure et de fonction qu'elles suscitent ? Pour répondre à cette question, nous devons déterminer quels sont les problèmes critiques, éviter les erreurs de catégorie et construire une théorie. Bien sûr, cette théorie doit être scientifique ; autrement dit, il doit être possible de la tester et de la réfuter par des moyens expérimentaux. Mais il n'est pas indispensable qu'elle permette *systématiquement* de faire des prédictions à tous les niveaux, ni qu'il soit immédiatement ou trivialement possible de réfuter chacun de ses éléments. (Si des critères aussi stricts avaient été appliqués à la théorie évolutionniste de Darwin au moment de sa conception, elle aurait été prématurément abandonnée.)

Au cours de la partie de ce livre qui suit, je présenterai un aperçu d'une telle théorie, que j'ai déjà longuement décrite dans ma trilogie sur la morphologie et l'esprit (voir la bibliographie). En fait, je ne présenterai que la quintessence de cet ensemble de livres. Mais je voudrais me simplifier la tâche en décrivant auparavant un certain nombre de systèmes biologiques connus qui possèdent des propriétés analogues à celles du cerveau. Néanmoins, je tiens à avertir le lecteur que ces analogies ne sont qu'heuristiques ; elles servent

1. Littéralement, « petite masse informe et visqueuse » (N.d.T.).

simplement à comprendre certains mécanismes, et ne doivent donc pas être prises comme exemples explicites des fonctions cognitives.

Avant d'aborder ces analogies, il peut être utile de réexaminer les problèmes dont nous sommes partis et de résumer le raisonnement que nous avons tenu jusqu'ici. Nous avons vu que, tant que la science et les observateurs scientifiques ne s'intéressaient qu'à des objets physiques et à des forces naturelles indépendants de l'esprit des observateurs, tout un ensemble de théories majeures, appartenant à un groupe cohérent de disciplines scientifiques, pouvait se permettre d'ignorer la complexité psychologique des observateurs scientifiques. Les chercheurs faisaient appel à leurs sensations et à leurs perceptions pour réaliser leurs expériences et pour communiquer de façon intersubjective avec leurs collègues, mais toute sensation et toute perception étaient bannies de leurs explications théoriques et formelles. Hormis quelques difficultés rencontrées aux frontières de l'extrêmement petit (dans les mesures quantiques) ou aux confins des très grandes vitesses ou des très grands objets (en physique relativiste), l'observateur scientifique *semblait* pouvoir tout à fait regarder les choses de l'extérieur. Il en est résulté une vision « objectiviste » de la nature, qui distinguait les choses les unes des autres grâce à des « catégories classiques » – des catégories définies par des conditions nécessaires et suffisantes. On établissait ensuite une correspondance dépourvue d'ambiguïté entre ces catégories et le monde physique en incorporant des données expérimentales dans des théories physiques d'une portée considérable.

Dans de nombreux domaines, cette approche fonctionnait (et fonctionne encore) à merveille, mais lorsque les études de physiologie et de psychologie du XIXe siècle ont redonné à l'esprit une place dans la nature, toute une série de difficultés ont commencé à émerger. Tout d'abord, l'observateur ne pouvait plus se permettre de négliger les phénomènes mentaux ni l'expérience mentale. Il ne pouvait plus ignorer la conscience elle-même ou l'intentionnalité de l'expérience consciente – autrement dit, le fait que celle-ci fasse toujours référence à un objet. Mais les mécanismes de cette conscience n'étaient pas directement transparents, et la conscience ne pouvait pas non plus être directement étudiée comme un objet externe – au mieux, elle pouvait être analysée par introspection ou, indirectement, à partir des comportements d'autrui.

Face à cet état de fait, l'une des réactions possibles consistait à déclarer le sujet hors limites. On insistait sur le fait que la science

devait s'intéresser seulement aux comportements qui peuvent être observés par les méthodes confirmées de la recherche scientifique ne traitant que d'objets dépourvus d'intentionnalité. Par la suite, pour tenter de sauvegarder l'attitude « scientifique » sans renier l'intentionnalité, et par opposition à cette attitude behavioriste, les sciences cognitives ont adopté un point de vue différent. Il consistait à s'approprier les concepts dérivés de l'analyse logique et formelle, en mettant l'accent sur la syntaxe. Dans cette perspective, l'esprit, tout comme un ordinateur, serait organisé selon certaines règles et opérerait par représentations mentales. La signification, la dimension sémantique apparaîtrait grâce à la mise en correspondance de ces règles avec des événements ou objets pouvant être classés dans des catégories classiques. Contrairement au behaviorisme, ce point de vue permettait d'étudier l'esprit. Cependant, il consistait à décrire le cerveau comme un système formel, c'est-à-dire d'une façon plus ou moins indépendante de sa structure détaillée. La correspondance sémantique établie entre cette description et le monde est objectiviste : les choses et les événements sont décrits, de façon univoque, comme des catégories classiques.

Comme je l'explique en détail dans la postface, l'idée que le cerveau et l'esprit fonctionnent comme des ordinateurs numériques ne résiste pas à l'analyse. Quant à l'idée qu'il existe des représentations mentales indépendantes de toute référence à des structures ou à des mécanismes cérébraux, son sort n'est guère meilleur. Il suffit de songer à la façon dont les animaux et les gens découpent le monde en catégories, et à celle dont les enfants se développent mentalement, pour s'apercevoir que le langage ne peut être correctement expliqué en termes purement syntaxiques, sans faire appel à une explication sémantique adéquate. La vision objectiviste du monde est au mieux incomplète et au pire complètement fausse. Le cerveau n'est pas un ordinateur et le monde n'est pas un morceau de bande magnétique.

Quand j'étais jeune chercheur, je croyais que la physique *parviendrait*, du moins un jour, à tout expliquer exhaustivement. J'étais objectiviste sans le savoir. Aujourd'hui, mon respect pour la physique est toujours aussi grand, mais je me suis rendu compte qu'il faut ajouter l'intentionnalité dans le tableau. L'histoire suivante illustre très bien ma vision actuelle de la structure des choses.

Un monsieur un peu paranoïaque était convaincu que sa petite amie le trompait avec un autre homme. Par une très chaude nuit d'été, il rentra à la maison plus tôt que de coutume et, en proie à

une crise de jalousie furieuse, il fouilla l'appartement de fond en comble à la recherche de l'hypothétique amant. En vain. Toujours aussi furieux, il se précipita dans la pièce du fond pour vérifier par la fenêtre si personne ne s'enfuyait par l'escalier de secours extérieur. Il aperçut alors un homme en train de défaire son col de chemise et de s'essuyer le front. Hors de lui, il attrapa son énorme réfrigérateur, le souleva et, après avoir bien visé, le laissa choir sur la tête de l'homme – suite à quoi il tomba raide mort d'épuisement.

La scène suivante se passe au Paradis. Trois hommes sont sur le point d'être admis ; saint Pierre leur dit qu'ils ont rempli toutes les formalités d'entrée sauf une, qui consiste à décrire les circonstances de leur mort. Le premier homme dit alors : « Eh bien, je pensais qu'il se passait des choses louches à la maison, et je suis donc rentré plus tôt. J'ai fouillé partout et j'ai fini par trouver ce type. Alors, j'ai dû avoir une brusque montée d'adrénaline. J'ai soulevé un frigidaire qui d'habitude aurait été trop lourd pour moi, je l'ai laissé tomber sur sa tête, et puis j'ai dû avoir une crise cardiaque. » Le deuxième homme dit : « Je n'en sais rien. C'était par une chaude nuit d'été. J'ai enjambé ma fenêtre pour passer sur l'escalier de secours, j'ai desserré le col de ma chemise, je me suis essuyé le front, et j'ai pris un frigidaire sur la tête. » Et le troisième homme dit : « Je n'en sais rien. J'étais assis dans un frigidaire, et je ne faisais de mal à personne. »

La physique des corps en chute libre ou celle de l'intentionnalité, sous-tendue par la morphologie de l'esprit, ne peuvent être perturbées par des idées fantaisistes non vérifiées ou des objets pesants sans graves conséquences.

Vouloir réfléchir à la question de l'esprit en l'absence de toute référence à la structure, à la fonction, au développement et à l'évolution du cerveau est une aventure intellectuelle risquée. Il est peu vraisemblable que l'on parvienne à deviner comment fonctionne le cerveau sans examiner sa structure. Et bien évidemment, si l'on est d'accord avec les éthologues pour dire que les états mentaux sont le produit d'une évolution, on doit au moins étudier la manière dont le cerveau a évolué. Nous sommes obligés d'achever le programme de Darwin.

Mais dès que nous commençons, d'un pas incertain, à avancer dans cette direction, nous nous retrouvons devant toute une série de résultats fascinants et en même temps déconcertants. Le développement du cerveau est radicalement dynamique et statistique. L'analyse de son développement nous incite à penser que les gènes

régissent l'anatomie cérébrale à la faveur d'interactions épigénétiques – parce que certains événements doivent survenir avant d'autres. Certaines molécules d'adhérence agissent sur des collectifs de cellules et sur leurs migrations, mais elles ne le font pas cellule par cellule, dans un ordre prescrit, prédéterminé. Et dans une certaine mesure, la migration et la mort des cellules sont des processus stochastiques – autrement dit, leurs conséquences au niveau de chaque cellule sont imprévisibles. Ces processus statistiques expliquent que chaque cerveau est unique, contrairement à ce qui se passe avec les ordinateurs. La diversité somatique nécessairement engendrée par ces moyens est si grande qu'elle ne peut pas être considérée simplement comme du « bruit » semblable au bruit qui survient dans les circuits électroniques aux températures de fonctionnement habituelles (la « friture » dans votre amplificateur, par exemple), et qui ne veut rien dire.

De fait, les circuits du cerveau ne ressemblent à aucun autre circuit connu. Les neurones possèdent des ramifications arborescentes qui se recouvrent partiellement et se divisent d'une myriade de façons différentes. Leur mode de communication n'est pas celui d'un ordinateur ou d'un central téléphonique ; l'émission de signaux dans le cerveau ressemble davantage à l'immense agrégat d'événements interactifs qui surviennent dans une forêt vierge. Et pourtant, malgré cela, les cerveaux donnent lieu à des cartes et à des circuits qui adaptent automatiquement leurs frontières aux modifications des signaux. Ils contiennent une multitude de cartes qui interagissent sans l'aide d'aucun superviseur, et qui garantissent l'unité et la cohésion des scènes perceptives. Et ils permettent aussi à leurs propriétaires (les pigeons, par exemple) de considérer comme faisant partie d'une même catégorie des ensembles très vastes – sinon infinis – d'objets très divers, tels les dessins de différents poissons, et ce après n'en avoir vu que quelques-uns.

Lorsque nous considérons à la fois ces propriétés cérébrales extraordinaires et les dilemmes engendrés par l'idée que le cerveau est une machine ou un ordinateur, nous pouvons dire à juste titre que nous nous trouvons devant une crise scientifique. La question se pose alors de savoir comment résoudre cette crise. Pour tenter de trouver une issue, tournons-nous vers la biologie elle-même plutôt que vers la physique, les mathématiques ou l'informatique.

TROISIÈME PARTIE

Propositions

Nous sommes à présent en mesure d'utiliser nos connaissances en biologie, en psychologie et en philosophie pour postuler une théorie de la conscience qui constituera un élément essentiel de la théorie du fonctionnement du cerveau.

La plupart des chercheurs considèrent qu'il est prématuré – et peut-être même insensé – de vouloir construire ce genre de théorie. Mais l'histoire des sciences incite à penser que, pour progresser, il ne suffit pas de rassembler des données ; il faut également produire des synthèses théoriques et les tester. Elle nous enseigne également que rien ne stimule davantage la réflexion et le travail expérimental que l'existence d'une théorie susceptible d'être amendée – et même démolie. Pour former une théorie scientifique, ses éléments doivent pouvoir être mis à l'épreuve, et elle doit contribuer à structurer tout – ou presque tout – ce que l'on sait sur le cerveau et sur l'esprit.

Pour bâtir une telle théorie à propos de la matière de l'esprit, il faudra passer en revue les divers niveaux d'organisation du système nerveux. Il faudra aussi repenser ce que nous entendons par « mémoire », « concepts », « signification », par « conscient comme un animal » et « conscient comme un être humain ». Au pire, cela permettra aux lecteurs de découvrir un certain nombre de phénomènes biologiques et de résultats fascinants. Au mieux, il leur sera ainsi possible d'entrevoir les bases matérielles de l'esprit. Je prie instamment les lecteurs d'être patients : nous abordons ici un lieu-frontière, un lieu aux limites mouvantes – un lieu où, malgré l'absence de points de références, notre curiosité s'exacerbe.

Chapitre 8

Les sciences de la reconnaissance

La sélection vaut mieux que les instructions.
Anonyme

À ce point de notre parcours, nous sommes prêts à aborder la question de l'esprit d'un point de vue biologique. Il s'agit là davantage d'une attitude intellectuelle – on considère les phénomènes mentaux en s'appuyant sur des concepts biologiques – que d'une démarche du type de ce que l'on pratique en biochimie, en biologie cellulaire ou en neurobiologie.

L'idée qu'il existe des modes de pensée caractéristiques de la biologie, indépendants des autres sciences, n'est pas communément admise. Pourtant, l'un de ces modes – et l'un des plus fondamentaux – a été largement développé par Darwin : il consiste à penser en termes de populations. Cela revient à considérer que les variations ne sont pas des erreurs, mais que, comme disait Ernst Mayr – l'un des grands noms de l'évolutionnisme –, elles sont réelles. La variabilité individuelle au sein d'une population constitue une source de diversité sur laquelle la sélection naturelle agit pour produire différentes sortes d'organismes. Ce point de vue est totalement opposé à l'essentialisme de Platon, qui requiert la création d'une typologie construite du haut vers le bas ; en effet, celui qui pense en termes de populations postule que l'évolution engendre des classes de formes vivantes, du bas vers le haut, par des processus sélectifs graduels s'étendant sur des milliards d'années (voir la figure 5-2). Ce type d'idées est étranger à la physique, car, même pour expliquer l'« évolution » des étoiles, on n'a besoin d'aucune notion de ce genre.

Pour relier les notions de biologie des populations aux idées sur l'esprit, il me faut détailler quelques-unes des implications de cette façon de penser, ainsi que celles du processus d'évolution proprement dit. Pour cela, je vais faire appel à un certain nombre de termes qui sont probablement nouveaux pour le lecteur, du moins dans leur acception spécialisée. Parmi eux se trouvent des termes

comme « instruction », « sélection », « reconnaissance », « mémoire » ou « héritabilité ». Remarquez que, à l'exception du dernier, tous ces mots sont d'usage courant ; mais ici, je vais les utiliser dans un sens quelque peu différent.

Commençons par le concept technique de « reconnaissance ». Je vais partir d'un énoncé abstrait, que je traduirai ensuite en un exemple issu de l'évolution. Par « reconnaissance », j'entends la mise en correspondance, adaptative et continuelle, des éléments d'un domaine physique donné aux nouveautés survenant dans les éléments d'un autre domaine physique plus ou moins indépendant du premier, ajustement qui a lieu en l'absence de toute instruction préalable. Difficile à avaler d'un seul coup ? Pour découper cette difficulté en morceaux plus digestes, permettez-moi de prendre l'évolution comme exemple.

Au cours de l'évolution, les organismes (les éléments du premier domaine) s'adaptent plus ou moins aux événements qui surviennent dans l'environnement (les éléments du second domaine). Cette adaptation a lieu même lorsque les modifications de l'environnement sont imprévisibles (c'est-à-dire lorsqu'elles constituent des nouveautés). Le processus d'adaptation réside dans la sélection des organismes présentant les variantes qui s'avèrent les plus aptes en moyenne, et ce qui les rend tels ne provient d'aucune information explicite *a priori* (ou « instruction ») sur la nature des nouveautés survenant dans l'environnement. Les modifications sélectives de l'environnement sont, en général, indépendantes de la variabilité existant au sein de la population d'organismes, bien que la sélection résultant de ces modifications puisse contribuer à la variabilité. En somme, aucun transfert explicite d'information entre l'environnement et les organismes n'est là pour amener la population à se modifier de façon à accroître son niveau d'adaptation. L'évolution opère par sélection, et non par instructions. Il n'y a pas de causes finales, pas de téléologie guidant le processus global ; les réponses surviennent *a posteriori* dans chaque cas.

L'idée est étonnante. Elle me rappelle cette dame, dans un livre de E. M. Forster, qui disait : « Comment pourrais-je savoir ce que je pense avant d'avoir vu ce que je dis ? » Plus étonnant encore est le fait que l'évolution, en agissant par sélection sur des populations d'individus et sur de longues périodes de temps, engendre des systèmes sélectifs *à l'intérieur* des individus. De tels systèmes sélectifs, qui agissent chez un organisme au cours de sa vie, forment ce que l'on appelle des systèmes sélectifs somatiques. Par conséquent, c'est

un système sélectif évolutif qui sélectionne un système sélectif somatique !

Je vais à présent décrire un exemple particulier, que l'on ne trouve que chez les vertébrés. Il s'agit du système immunitaire. Le fait d'avoir compris les principes de base de l'immunité nous sera très utile pour comprendre la sélection dans le système nerveux. Cette description justifiera, je l'espère, la façon plutôt spécialisée et abstraite dont j'utilise le mot « reconnaissance ». Le système immunitaire est l'exemple le mieux connu de système somatique de reconnaissance fondé sur des principes de sélection. En fait, je me propose de démontrer ici qu'il existe des sciences de la reconnaissance, c'est-à-dire des sciences qui étudient les systèmes de reconnaissance. Les données abondent qui incitent à penser que l'évolution et l'immunité adaptative sont deux systèmes sélectifs agissant sur des populations à des échelles de temps différentes. Tout cela pour dire qu'il semble bien que le cerveau se comporte lui aussi comme un système sélectif somatique, et donc que la neurobiologie fait elle aussi partie des sciences de la reconnaissance.

Mais je vais trop vite. Je voudrais plutôt commencer par décrire brièvement le système immunitaire, à l'étude duquel j'ai consacré environ quinze années de ma vie, car c'est à la fois fascinant et instructif. Le système immunitaire de chaque individu est un système sélectif somatique composé de molécules, de cellules et d'organes spécialisés. En tant que système, il est capable de distinguer au niveau moléculaire entre soi et non-soi. Par exemple, il est chargé de détecter les caractéristiques chimiques des envahisseurs bactériens et viraux (le non-soi) et d'y réagir. Autrement, ces envahisseurs viendraient à bout des ensembles de systèmes cellulaires des individus (le soi). La réaction du système immunitaire fait appel à une reconnaissance moléculaire incroyablement spécifique. Un système immunitaire convenablement stimulé est capable de faire la différence entre deux grandes molécules protéiques étrangères composées chacune de milliers d'atomes de carbone et ne se distinguant l'une de l'autre que par l'orientation légèrement différente (quelques degrés) d'une seule de leurs chaînes carbonées. Et il peut également distinguer ces molécules de toutes les autres molécules et, ayant initialement acquis la capacité de le faire, conserver cette capacité. Autrement dit, il possède une « mémoire ».

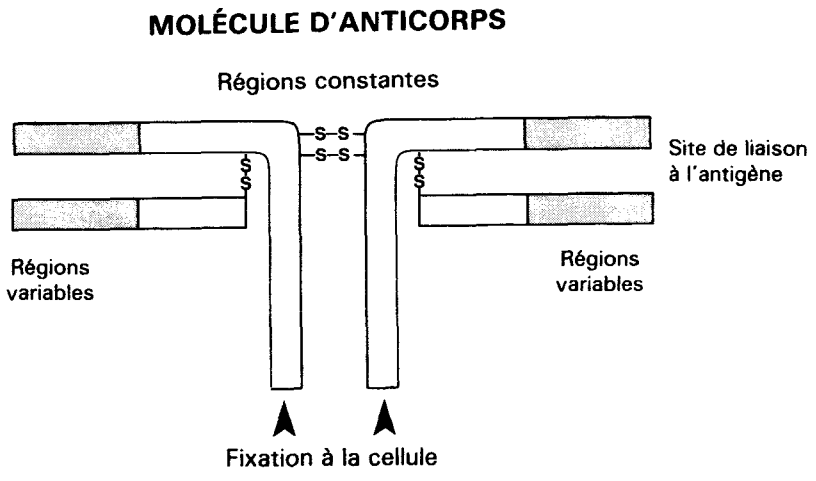
Si j'injecte dans l'organisme d'un individu une protéine qui ne ressemble pas à ses propres protéines, certaines cellules spécialisées, les lymphocytes, vont réagir en fabriquant des molécules appelées

anticorps (figure 8-1). Celles-ci vont se lier à l'intrus en s'emboîtant sur des portions spécifiques et caractéristiques de cette molécule, appellée antigène. Par la suite, lors d'une deuxième rencontre, ces anticorps ne se lieront qu'à ces antigènes-là, de façon plus efficace encore. Le plus étonnant est peut-être le fait qu'une reconnaissance spécifique ait lieu même vis-à-vis des nouvelles molécules synthétisées par les chimistes organiciens – des molécules qui n'ont jamais existé auparavant, ni chez l'espèce produisant la réaction immunitaire, ni sur Terre d'ailleurs.

Comment se fait-il que le corps d'un individu soit capable de reconnaître formellement des molécules inédites d'une manière aussi spécifique ? La théorie qui prévalait avant l'actuelle s'appelait théorie de l'instruction. L'hypothèse de base de cette théorie était que, dans le système immunitaire, les molécules étrangères transfèrent de l'information concernant leur forme et leur structure au site de liaison des molécules d'anticorps. Ensuite, lorsqu'elles se retirent de ce site (telles des formes à découper des biscuits dans la pâte), elles laissent une empreinte dont la forme est complémentaire de la leur et qui pourra donc par la suite se lier à toutes les molécules étrangères possédant des régions dont la forme sera identique à celle qui a initialement laissé son empreinte. À l'évidence, il s'agit là d'un processus instructif : on postule qu'il est *nécessaire* de transmettre de l'information à propos d'une structure tridimensionnelle afin d'instruire le système immunitaire quant à la façon de construire la molécule d'anticorps dont la chaîne polypeptidique, en s'enroulant autour de cette structure (figure 6-1, en bas), créera la forme complémentaire adéquate.

Il se trouve que cette théorie simple et élégante est fausse. Celle qui l'a remplacée est plus complexe, et même contre-intuitive, mais elle permet d'expliquer un large éventail de données ; de fait, presque aucun immunologiste aujourd'hui ne remet en cause sa validité essentielle. Cette théorie est connue sous le nom de théorie de la sélection clonale ; elle a été formulée pour la première fois par Sir Frank Macfarlane Burnet, aujourd'hui décédé.

Selon Burnet, avant même une confrontation avec des molécules étrangères, l'organisme de tout individu est capable de fabriquer un immense répertoire de molécules d'anticorps ayant chacune un site de liaison de forme différente. Lorsqu'une molécule étrangère (présente sur un virus ou une bactérie, par exemple) s'introduit dans le corps, elle rencontre une *population* de cellules portant



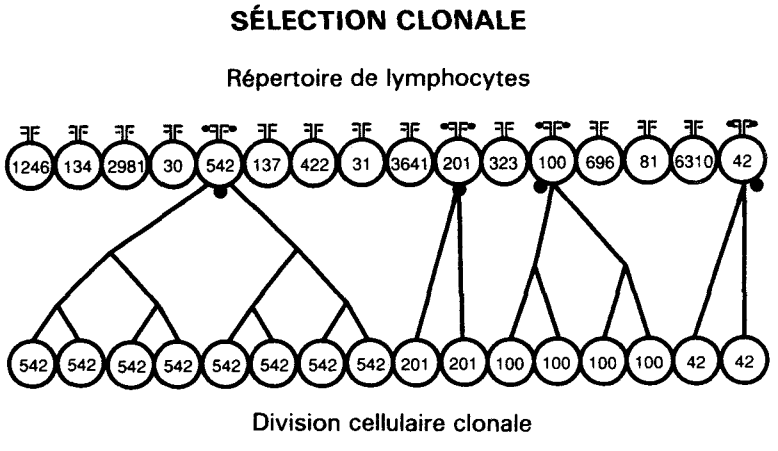


FIGURE 8-1 : *Le système immunitaire fonctionne comme un système sélectif de reconnaissance. Votre système immunitaire sait distinguer les molécules étrangères (non-soi) des molécules de votre corps (soi) en vertu du fait qu'elles ont des formes différentes. Il y parvient en fabriquant des protéines appelées anticorps. Chaque cellule immunitaire fabrique un anticorps dont la région variable est différente de celle des autres* (en haut) : *chaque région variable possède un site de liaison dont la forme est différente. Lorsqu'une molécule étrangère ou antigène (représentée par les points noirs* sur le dessin du bas*) pénètre dans l'organisme, elle ne se lie qu'aux anticorps (présents à la surface des cellules immunitaires) dont il se trouve que la forme s'ajuste à celle de certaines de ses parties (cellules 542, 201, 100 et 42). Cet ensemble de cellules se divise alors et forme des " clones " – des populations de cellules semblables, portant chacune des anticorps semblables. Lorsque l'antigène sera présenté une seconde fois, un grand nombre de copies de ces mêmes anticorps seront là pour aider à le détruire. Les cellules numérotées 542, 201, 100 et 42, par exemple, seront plus nombreuses que les autres, et reconnaîtront plus rapidement les molécules étrangères la prochaine fois que celles-ci s'introduiront dans l'organisme. Les intruses peuvent être des molécules présentes à la surface d'un virus ou d'une bactérie. Ce système est sélectif parce qu'un grand nombre de formes d'anticorps différentes, susceptibles de se lier à des antigènes, existe (chacune sur une cellule différente)* avant *l'entrée en scène des antigènes. Les antigènes ne sélectionnent qu'un petit nombre de ces formes d'anticorps, dont la production est alors énormément amplifiée par la division clonale des cellules correspondantes (2,4,8,16,...), ce qui aboutit à la présence d'énormes quantités de ces anticorps-là. Ainsi, la population d'anticorps se modifie avec l'expérience.*

chacune un anticorps *différent* à sa surface. La molécule étrangère se lie alors aux cellules du répertoire qui portent des anticorps dont le site de liaison se trouve être plus ou moins complémentaire de sa propre forme. Or, dès qu'une portion de l'antigène s'emboîte suffisamment bien sur un anticorps, cela incite la cellule (appelée lymphocyte) portant cet anticorps à se diviser de nombreuses fois (figure 8-1). Cela donne naissance à une « progéniture » plus nombreuse de cellules portant des anticorps de même forme et de même spécificité de liaison.

Les groupes de cellules-filles sont des clones (on appelle clone la progéniture par reproduction asexuée d'une unique cellule), et l'ensemble du processus consiste en une reproduction différentielle par sélection clonale. En d'autres termes, à l'issue de la sélection des cellules qui portent les anticorps spécifiques adéquats au sein d'un grand répertoire de cellules différentes, le nombre d'anticorps dont le type est spécifique d'une forme étrangère particulière s'accroît parce que la liaison sélective a induit la multiplication *a posteriori* des cellules portant ces anticorps-là. La composition de la population de lymphocytes a donc été modifiée par sélection.

L'analyse de la structure complète des anticorps a été réalisée dans mon laboratoire il y a déjà plusieurs décennies. Elle a permis de montrer que les chaînes polypeptidiques des anticorps (figure 8-1, en haut) se composent de régions constantes (semblables ou identiques chez toutes ces molécules) et de régions variables (qui diffèrent selon le type de molécule et contiennent le site de liaison). Nous savons aujourd'hui que cette diversité est engendrée somatiquement (c'est-à-dire au cours d'une vie humaine) dans les lymphocytes de l'organisme de chaque individu. Ce processus fait appel à une sorte de brassage, au sein de chaque lymphocyte, du code génétique spécifiant les régions variables des anticorps *susceptibles* de se lier un jour à un antigène.

J'espère en avoir dit assez pour vous montrer que le système immunitaire correspond bien à ma définition des systèmes de reconnaissance. Il existe dans un domaine physique donné (le corps d'un individu) et il réagit aux nouveautés qui émergent de façon indépendante dans un autre domaine (à certaines molécules étrangères parmi les millions de millions de molécules chimiquement différentes possibles) par l'intermédiaire d'une liaison spécifique et d'une réponse cellulaire adaptative. Il fait cela *sans* qu'aucune information concernant la forme à reconnaître n'ait besoin d'être transmise au système de reconnaissance *au moment où celui-ci fabrique les molé-*

cules qui opéreront la reconnaissance, c'est-à-dire les molécules d'anticorps. Au contraire, le système de reconnaissance engendre *d'abord* une population diversifiée de molécules d'anticorps pour ensuite sélectionner *a posteriori* celles qui s'ajustent le mieux. Il fait cela continuellement et, la plupart du temps, de façon adaptative.

Le système sélectif immunitaire possède un certain nombre de propriétés fascinantes. Premièrement, il existe plus d'une façon de réussir à reconnaître une forme donnée. Deuxièmement, cette façon n'est jamais exactement la même d'un individu à un autre ; autrement dit, deux individus donnés n'ont jamais des anticorps identiques. Troisièmement, le système possède une sorte de mémoire cellulaire. Suite à la présentation de l'antigène à un ensemble de lymphocytes capables de s'y lier, certains de ces lymphocytes ne se diviseront que peu de fois, tandis que les autres s'engageront de façon irréversible dans la production d'anticorps spécifiques de l'antigène, puis mourront. De ce fait, les quelques cellules qui se sont divisées sans aller jusqu'à la production d'anticorps forment alors, au sein de la population cellulaire totale, un groupe de cellules plus grand qu'il ne l'était initialement. Par la suite, ce groupe pourra réagir au même antigène de façon beaucoup plus rapide. Comme je l'ai déjà évoqué, cela veut donc dire que le système présente une sorte de mémoire au niveau cellulaire.

Voilà donc un système de reconnaissance moléculaire qui est non cognitif et hautement spécifique. Son fonctionnement s'explique à merveille par un raisonnement en termes de populations, qui correspond à l'essence même du darwinisme. Tout comme l'évolution, ce système possède un générateur de diversité (le « brasseur » d'ADN dans chaque lymphocyte), un moyen de perpétuer les modifications grâce à une sorte d'hérédité (la division clonale), et un moyen d'amplifier différentiellement les événements sélectifs (la reproduction différentielle des clones). Mais, contrairement à l'évolution, il a lieu dans des *cellules*, il opère sur de *courts laps de temps* et il n'engendre pas de nombreux niveaux morphologiques, mais seulement des molécules d'anticorps différentes. Il s'agit d'un système sélectif somatique.

Remarquez qu'au cours de l'évolution elle-même, la diversité est engendrée au sein des populations *animales* par des mutations survenant dans l'ADN. Celles-ci sont transmises héréditairement par l'intermédiaire des cellules germinales (les spermatozoïdes et les ovocytes). La sélection opère ensuite continuellement sur des

individus, à l'échelle de temps de *l'évolution*, pour donner naissance à différentes espèces, en fonction d'un grand nombre de variables présentes dans l'environnement. Les deux systèmes – évolution et immunité – font donc appel à des principes de sélection semblables pour traiter les nouveautés, mais leurs mécanismes d'action sont très différents. Du point de vue conceptuel, il est important de faire la part entre le principe sélectif et les mécanismes mis en œuvre pour l'exprimer dans chaque système physique particulier.

Ce que ces deux sciences de la reconnaissance, l'évolution et l'immunologie, ont en commun, ne se rencontre pas dans les systèmes non biologiques tels que les étoiles « en cours d'évolution ». Les systèmes physiques de ce genre peuvent s'expliquer en termes de transferts d'énergie, de dynamique, de causes et même de « transferts d'information ». En revanche, ils ne possèdent aucun répertoire de variantes prêtes à interagir par sélection pour fournir une réponse, en tant que population, selon un principe héréditaire. À ce propos, l'application d'un principe sélectif à un système de reconnaissance n'implique pas *nécessairement* la participation des gènes – elle signifie simplement que tout état issu d'une sélection sera fortement corrélé, du point de vue structurel, avec celui qui lui a donné naissance, et que la corrélation continuera à se propager. Il n'est pas non plus exact que la sélection n'est pas elle-même capable d'introduire de la variabilité. Néanmoins, il *faut* qu'il y ait persistance, autrement dit qu'il subsiste une « mémoire », des événements sélectifs. Si les modifications survenaient trop vite pour que les caractéristiques sélectionnées puissent émerger dans la population et se maintenir, les systèmes de reconnaissance ne survivraient pas. Quoi qu'il en soit, la physique proprement dite ne traite pas des systèmes de reconnaissance, qui sont par nature des systèmes biologiques et historiques. Mais toutes les lois de la physique valent pour les systèmes de reconnaissance.

Leo Szilard, un grand physicien dont les expériences menées en collaboration avec Enrico Fermi ont abouti à la fission nucléaire contrôlée, était fasciné à la fois par l'immunologie et le cerveau. Il venait souvent dans mon laboratoire pour voir ce qu'il y avait de neuf sur les anticorps. D'habitude, il commençait par dire : « Eh bien ! Quel est le problème ? Je n'ai qu'un quart d'heure. » Une fois, Szilard a participé à un congrès au cours duquel un chercheur eut le malheur de proposer une théorie de la mémoire. La pensée, expliqua ce chercheur, induisait la fabrication de nouvelles protéines

par notre cerveau. Passé un certain temps, ces protéines nouvelles en stimulaient d'autres, semblables aux anticorps, qui représentaient des souvenirs. Leo se leva alors et dit, en arborant un sourire joyeux et impitoyable : « C'est peut-être comme ça que *votre* cerveau fonctionne. »

Le cerveau constitue-t-il un système sélectif de reconnaissance ? La description des opérations de base du cerveau en ces termes se révélera-t-elle fructueuse et utile ? Comme vous l'avez sans doute déjà compris, je pense qu'elle sera non seulement utile, mais qu'elle permettra également d'éliminer une grande part du paradoxe et de la crise auxquels on est confronté lorsqu'on étudie les données existant sur la structure et la fonction cérébrales. En fait, je suis convaincu que la neurobiologie est une science de la reconnaissance. Mais, malgré l'excitation ressentie par mes collègues et moi-même lorsqu'à la suite de nos travaux sur les anticorps, nous avons découvert que les molécules d'adhérence des cellules nerveuses – la « colle du cerveau » – sont les précurseurs, du point de vue évolutif, de l'ensemble du système immunitaire, je n'ai pas l'intention d'imiter l'infortuné interlocuteur de Szilard. La similitude entre le système immunitaire et le système nerveux n'existe qu'au niveau des principes, et non au niveau des détails.

J'ai défendu l'idée selon laquelle le cerveau est un système sélectif de reconnaissance parce que le fait de réfléchir à la fonction cérébrale en termes sélectifs nous épargne les horreurs de l'homoncule (figure 8-2). Puisque la diversité existe au préalable dans les systèmes sélectifs, la spécificité résultant d'une sélection *a posteriori*, nous n'avons plus besoin de supposer l'existence, dans le cerveau, d'une suite sans fin de processeurs d'information emboîtés les uns dans les autres. Et pour justifier cette affirmation, je vais décrire une théorie de la fonction cérébrale qui suit des principes sélectifs. La difficulté consiste à montrer comment l'évolution et le développement donnent naissance à un système somatique de sélection dans le cerveau. Ensuite, j'essayerai de montrer que les mécanismes de sélection proposés peuvent rendre compte des fonctions psychologiques – de la perception, de la mémoire, et même de la conscience. Attelons-nous à présent à cette tâche.

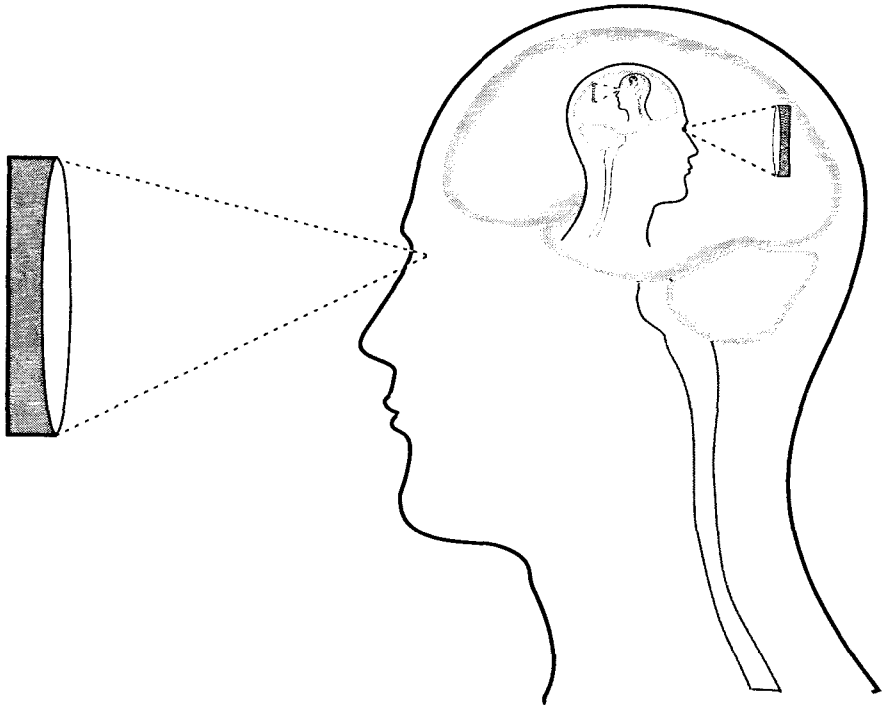


FIGURE 8-2

La suite sans fin des homoncules emboîtés. La notion d'instruction ou de traitement de l'information exige qu'il y ait quelqu'un ou quelque chose pour lire l'information. Mais une entité semblable s'avère alors nécessaire pour lire les messages résultants, et ainsi de suite à l'infini.

Chapitre 9

Le darwinisme neuronal

S'il est question de donner un nom à tout cet ensemble d'idées, je suggère qu'on l'appelle « edelmanisme neuronal ».

Francis H. C. Crick

Si nous considérons la reconnaissance comme une sorte d'ajustement adaptatif, la raison pour laquelle elle s'applique à l'évolution et à l'immunité devient évidente. Dans les deux cas, raisonner sur des populations est opératoire. Mais qu'est-ce qui justifie que l'on raisonne en ces termes pour étudier les mécanismes cérébraux – qu'est-ce qui justifie le darwinisme neuronal ? Il n'est pas opportun d'entrer ici dans toutes les complexités d'un tel point de vue, mais je crois que le fait de présenter certaines des raisons qui mènent à considérer la neurobiologie comme une science de la reconnaissance permettra d'éclaircir une grande part de ce que j'aurai à dire dans la suite de ce livre.

La première raison est presque trop évidente : la neurobiologie et l'étude du comportement s'intéressent à l'ajustement adaptatif des animaux à leur environnement. En considérant la neurobiologie comme une science de la reconnaissance, je laisse donc entendre que la reconnaissance n'est pas un processus qui procède par instructions. Il n'y a aucun transfert direct d'information, tout comme il n'y en a aucun dans les processus évolutifs ou immunitaires. Au contraire, la reconnaissance est sélective.

On trouvera plusieurs justifications de ce point de vue dans les critiques que j'ai formulées plus haut (chapitre 3) à propos des diverses erreurs de catégorie qui peuvent être commises lorsqu'on réfléchit au cerveau ; on trouvera également des arguments plus détaillés en faveur de ces critiques dans la postface. J'ai déjà avancé l'idée que le monde n'est pas un morceau de bande magnétique et que le cerveau n'est pas un ordinateur. Cependant, si nous adoptons mon point de vue, nous devrons montrer comment des animaux qui présentent des comportements définis parviennent néanmoins

à ajuster leurs réactions de façon adaptative aux nouveautés imprévues surgissant dans un tel monde.

Il existe également un autre ensemble de raisons incitant à penser que la reconnaissance ne peut dépendre d'instructions. Nous avons déjà vu, en effet, que les modèles supposant que le cerveau est un ordinateur ont du mal à rendre compte de l'individualité et de la diversité structurelle des différents cerveaux, même au sein d'une espèce donnée. En fait, les données issues de l'embryologie incitent à penser que l'extraordinaire diversité anatomique présente au niveau des ramifications les plus fines des réseaux neuronaux est l'une des conséquences inévitables du processus de développement. Un tel degré de diversité individuelle ne peut être toléré dans un ordinateur qui suit des instructions. Mais c'est exactement ce dont a besoin un système sélectif.

Une autre raison, concernant cette fois-ci l'homoncule, incite elle aussi fortement à adopter un point de vue sélectif plutôt qu'instructif. Je rappelle que l'homoncule est ce personnage dont on doit postuler l'existence « au sommet de l'esprit » dans toute théorie fondée sur des instructions, et qui agit comme interprète des signaux et des symboles. Si l'on considère que l'information provenant du monde extérieur est traitée selon certaines règles dans un cerveau semblable à un ordinateur, l'existence de l'homoncule paraît s'imposer. Mais alors, il faut aussi un second homoncule dans *sa* tête à lui, et ainsi de suite à l'infini (figure 8-2). En revanche, les systèmes sélectifs, dans lesquels l'ajustement s'opère *a posteriori* sur un répertoire diversifié *déjà existant*, ne requièrent aucune création particulière, aucun homoncule, aucune suite infinie emboîtée.

Si nous admettons que les fonctions cérébrales sont bâties selon un processus sélectif, nous devons alors être capables de réconcilier la variabilité structurelle et fonctionnelle du cerveau avec la nécessité d'expliquer comment il effectue des catégorisations. Et pour ce faire, nous avons besoin d'une théorie qui possède un certain nombre de caractéristiques essentielles : elle doit s'accorder avec les données de l'évolution et du développement ; elle doit rendre compte de la nature adaptative des réponses à la nouveauté ; elle doit permettre de montrer comment les fonctions du cerveau se développent proportionnellement à celles du corps à mesure que celui-ci se modifie sous l'effet de la croissance et de l'expérience ; elle doit rendre compte de l'existence de cartes dans le cerveau et des fonctions de ces cartes – pourquoi elles fluctuent, comment il se fait qu'une multiplicité de cartes puisse conduire à des réponses intégrées, et

comment, même en l'absence de langage, elles conduisent à des généralisations des réponses perceptives. Une telle théorie devrait également parvenir un jour à rendre compte de l'émergence du langage lui-même. Et enfin, elle doit permettre d'expliquer comment sont apparues, au cours de l'évolution, les différentes formes de catégorisation perceptive et conceptuelle, de mémoire et de conscience.

Pour être scientifique, la théorie doit partir de l'hypothèse que la cognition et l'expérience consciente ne reposent que sur des processus et des types d'organisation qui existent dans le monde physique. Elle doit par conséquent prendre soin de préciser quels sont les liens existant entre processus psychologiques et physiologiques.

La théorie que j'ai formulée afin de rendre compte de tous ces aspects est connue sous le nom de théorie de la sélection des groupes neuronaux (TSGN). Je vais en donner ici les principes de base, tout en mettant l'accent sur celles de ses caractéristiques qui permettent de construire un pont entre la psychologie et la physiologie. Si je tiens à expliquer cette théorie de façon un tant soit peu détaillée, c'est parce que ce faisant nous aborderons la catégorisation perceptive, les concepts, la mémoire et l'apprentissage, et surtout parce que cela nous permettra de nous attaquer au problème troublant de la conscience.

Je tiens à dire ici au lecteur qu'il va me falloir expliquer un certain nombre de processus complexes, car il faut les avoir compris pour comprendre la fonction cérébrale. Les principales notions à saisir sont ce qu'on appelle la sélection des groupes neuronaux, la « réentrée » et les cartographies globales. Je fournirai des exemples de chacune de ces notions. Une fois maîtrisées, elles seront très utiles, puisque nous nous en servirons maintes et maintes fois au cours des chapitres ultérieurs.

Dans la mesure où elle fait appel à tous ces processus, la TSGN est une théorie complexe ; ses principes de base, cependant, ne sont qu'au nombre de trois. En effet, même lorsqu'il s'agit d'expliquer une propriété aussi remarquable que la conscience, il n'en faut pas plus. Ce dont nous *avons* besoin, en revanche, pour expliquer une telle propriété, c'est de l'émergence, au cours de l'évolution, de nouvelles sortes de morphologies, à la fois dans le corps et dans le cerveau. Par conséquent, je décrirai au fur et à mesure un certain nombre de caractéristiques de ces morphologies.

Les trois principes de base de la TSGN (figure 9-1) concernent la façon dont l'anatomie du cerveau est initialement mise en place au cours du développement, la façon dont certains types de réponses sont ensuite sélectionnés dans cette anatomie à travers l'expérience, et la façon dont la réentrée – le processus par lequel les cartes cérébrales résultantes s'envoient des signaux – donne naissance à des fonctions importantes du point de vue du comportement.

Selon le premier principe, la sélection qui s'opère au cours du développement – c'est-à-dire les processus dynamiques primaires que nous avons abordés dans les chapitres 3 et 6 – conduit à la formation de la neuro-anatomie caractéristique d'une espèce donnée. Cette anatomie présente obligatoirement une variabilité énorme à ses niveaux les plus microscopiques et à ses ramifications les plus fines. C'est dû à la régulation dynamique opérée par les CAM et les SAM, aux fluctuations stochastiques des mouvements cellulaires, de la croissance des prolongements cellulaires et de la mort cellulaire au cours du développement, et également à l'ajustement des connexions neuronales, qui dépend de l'activité de neurones et qui se superpose au processus de croissance de leurs ramifications (ou neurites) à mesure que celles-ci explorent une région donnée du cerveau en développement. L'ensemble de ce processus est de nature sélective, et fait intervenir des populations de neurones engagées dans une *compétition* topobiologique. On appelle *répertoire primaire* toute population faite de groupes neuronaux différents appartenant à une région cérébrale donnée et comportant des réseaux neuronaux mis en place par des processus de sélection somatique. Le code génétique ne fournit pas de diagramme spécifique pour le câblage d'un tel répertoire. Il impose plutôt un ensemble de *contraintes* sur le processus de sélection. Cependant, malgré ces contraintes, même les individus génétiquement identiques sont peu susceptibles de posséder des câblages identiques, car la sélection est un phénomène épigénétique.

Le second principe de la TSGN fournit un mécanisme de sélection supplémentaire qui, en général, ne fait pas intervenir des modifications de la structure anatomique. Ce principe part de l'hypothèse selon laquelle, à cause des comportements de l'animal, les connexions synaptiques au sein de cette anatomie sont sélectivement renforcées ou affaiblies par des processus biochimiques spécifiques. Ce mécanisme, qui est sous-jacent à la mémoire ainsi qu'à un certain nombre d'autres fonctions, « taille » effectivement, par sélection, divers *circuits* fonctionnels (ceux dont les synapses ont été ren-

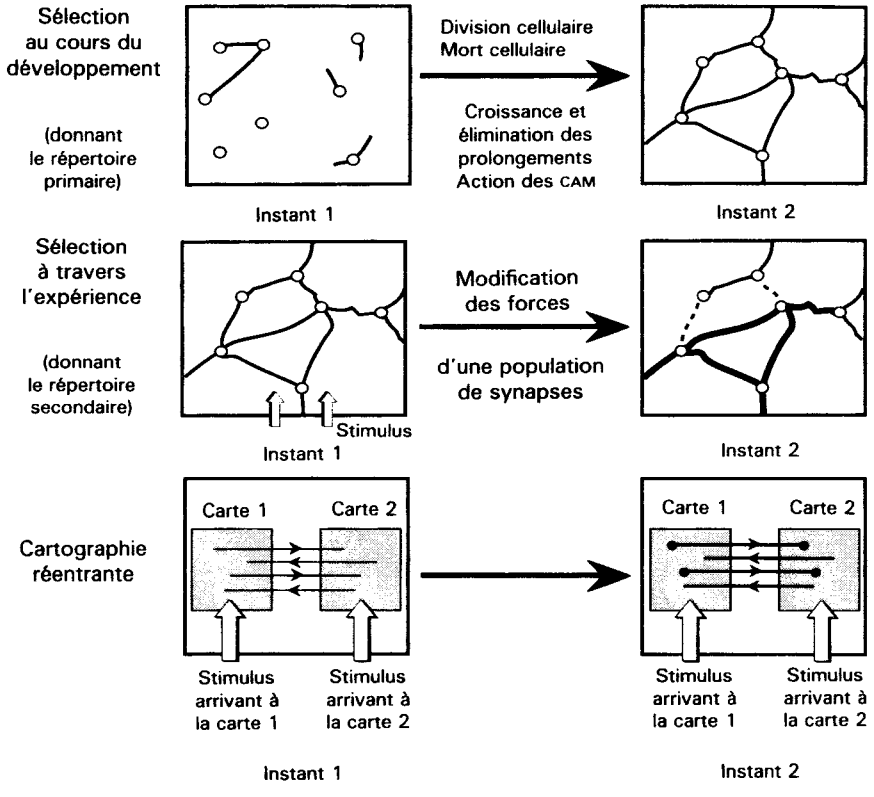


FIGURE 9-1

Une théorie sélectionniste de la fonction cérébrale. Appelée théorie de la sélection des groupes neuronaux, elle repose sur trois principes de base. En haut : *sélection au cours du développement. Celle-ci se produit en raison des effets moléculaires de la régulation due aux* CAM *et aux* SAM, *de l'émission de facteurs de croissance et de la mort sélective des cellules, et aboutit à la formation de réseaux anatomiques variés chez chaque individu. Ces réseaux constituent le* répertoire primaire. Au centre : *sélection à travers l'expérience. Les comportements induisent le renforcement ou l'affaiblissement sélectifs de diverses populations de synapses, ce qui conduit à la formation de divers circuits, constituant un* répertoire secondaire *de groupes neuronaux. Les conséquences des renforcements synaptiques sont indiquées par des traits gras ; celles des affaiblissements par des traits pointillés.* En bas *: la réentrée. Les liaisons entre cartes s'établissent au bout d'un certain temps à travers la sélection en parallèle et la mise en corrélation des groupes neuronaux des différentes cartes, qui reçoivent des entrées de façon indépendante et disjointe. Ce processus sert de base à la catégorisation perceptive. Les points visibles aux extrémités des connexions réciproques actives indiquent un renforcement en parallèle et plus ou moins simultané des synapses dans des circuits réentrants. (Les lecteurs qui souhaiteraient se rafraîchir la mémoire concernant les synapses peuvent se reporter à la figure 3-2.) Le renforcement (ou l'affaiblissement) peut se produire à la fois dans les connexions réentrantes intrinsèques et extrinsèques.*

forcées) dans le réseau anatomique. Un tel ensemble de circuits fonctionnels différents est appelé *répertoire secondaire.*

Dans une certaine mesure, les mécanismes qui aboutissent à la formation des répertoires primaires et secondaires s'entremêlent. C'est dû au fait que, à certains moments et à certains endroits – comme par exemple au cours de l'ajustement des connexions en fonction de l'activité neuronale (voir un exemple sur la figure 3-4) –, la formation du répertoire primaire dépend de la modification de la force des synapses. Même dans un cerveau entièrement développé, des « bourgeonnements » peuvent se produire, au cours desquels de nouveaux prolongements neuronaux peuvent former des synapses supplémentaires. Dans certains cas, tels le développement du chant chez les oiseaux ou la métamorphose chez la grenouille, la formation de nouvelles parties du système nerveux, avec formation simultanée de répertoires primaires et secondaires, se produit pendant que l'animal interagit avec le monde.

Le troisième principe de la TSGN concerne la façon dont les phénomènes sélectifs décrits dans les deux premiers agissent pour relier la psychologie à la physiologie. Il indique comment les cartes cérébrales interagissent grâce à un processus appelé réentrée. C'est peut-être la plus importante de toutes les propositions de la théorie, car elle sous-tend la façon dont les aires du cerveau qui apparaissent au cours de l'évolution se coordonnent pour donner naissance à de nouvelles fonctions.

Pour accomplir ces fonctions, les répertoires primaires et secondaires doivent former des cartes. Celles-ci sont reliées entre elles par des connexions réciproques et massivement parallèles. Le système visuel du singe, par exemple, possède plus de trente cartes différentes, chacune présentant un certain degré de ségrégation fonctionnelle (vis-à-vis de l'orientation, de la couleur, du mouvement, etc.) et reliées aux autres par des connexions parallèles et réciproques (figure 9-2). Des échanges de signaux réentrants ont lieu le long de ces connexions. Cela signifie que, à mesure que des groupes de neurones sont sélectionnés dans une carte, d'autres groupes, situés dans d'autres cartes – reliées à la première de façon réentrante –, pourront être sélectionnés en même temps. La mise en corrélation et la coordination de ces événements sélectifs sont obtenues par émission de signaux réentrants et par renforcement des interconnexions entre les cartes pendant un certain laps de temps.

L'une des prémisses fondamentales de la TSGN est que la coor-

dination sélective des structures complexes d'interconnexion entre groupes neuronaux par réentrée est à la base du comportement. En fait, la réentrée (alliée à la mémoire, que j'aborderai plus tard), forme la base du pont qui permettra de relier la physiologie à la psychologie.

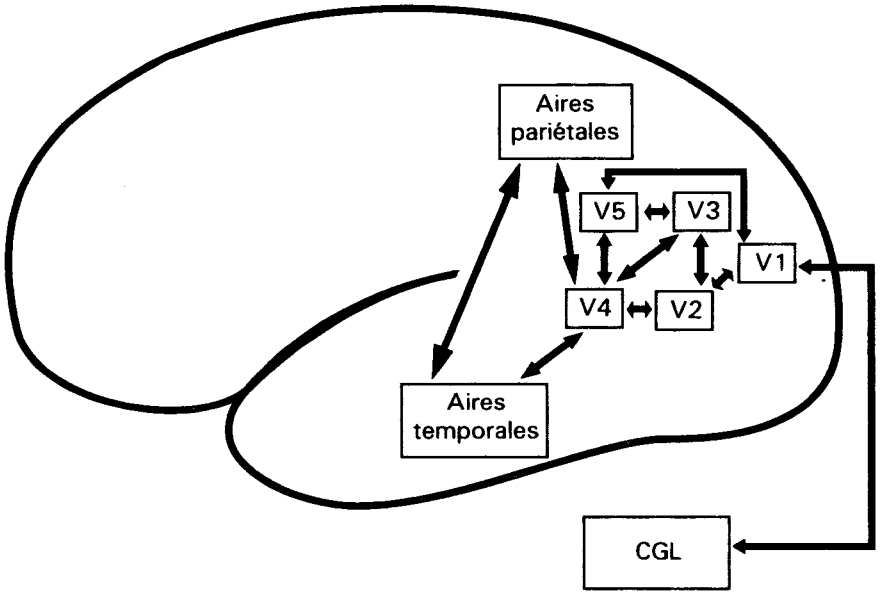


FIGURE 9-2

Les multiples cartes des aires visuelles du cerveau sont connectées entre elles de façon réentrante (voir les doubles flèches reliant les cartes visuelles V1-V5, les aires temporales et les aires pariétales). Il existe une ségrégation fonctionnelle entre les cartes – vis-à-vis de la couleur, du mouvement, de l'orientation et ainsi de suite. Mais il n'existe aucune « carte-en-chef » chargée de résumer « l'information » concernant ces propriétés. Cependant, grâce à la réentrée (les doubles flèches), les cartes réagissent de façon cohérente aux combinaisons de ces propriétés. Même la région connue sous le nom de corps genouillé latéral (CGL) – une région sous-corticale qui reçoit des signaux provenant du nerf optique de l'œil – est connectée de façon réentrante à la carte visuelle primaire, appelée V1.

Je n'ai pas encore précisé quelle était l'unité de sélection dans le cerveau. Pour ce qui concerne l'évolution, la principale unité sur laquelle opère la sélection est l'animal lui-même (le phénotype) ; pour l'immunité, c'est le lymphocyte. Selon la TSGN, cependant, l'unité de sélection dans le cerveau n'est pas la cellule nerveuse, mais plutôt un ensemble de cellules étroitement connectées, appelé groupe neuronal (figure 9-3). Le fait qu'il en soit ainsi est lié aux limitations des propriétés des neurones, aux contraintes du déve-

loppement et aux exigences neuro-anatomiques qui pèsent sur les circuits réentrants. En effet, chaque neurone est soit excitateur pour d'autres neurones, soit inhibiteur, mais pas les deux à la fois. En revanche, les groupes de neurones, qui comportent un mélange des deux sortes de neurones, peuvent être les deux à la fois. Au cours de la formation du répertoire primaire, les neurones voisins ont tendance à établir des connexions plus abondantes entre eux, formant ainsi des circuits qui contiennent des proportions variables de chaque sorte de neurone. De fait, cela confère un caractère hautement coopératif à l'activité des groupes de neurones, activité qui, selon toute vraisemblance, sera différente selon les aires et les cartes considérées, puisque les répertoires primaires respectifs le sont. En fait, il existe une raison encore plus convaincante de penser que les groupes neuronaux sont les unités de sélection. Lorsque des cartes sont reliées par des fibres réentrantes, chaque fibre étend en général ses arborisations sur de nombreux neurones localement reliés entre eux (figure 9-3). Et, lors de la formation des répertoires secondaires, le renforcement des synapses *au sein* de ces arborisations peut conduire à la sélection de groupes de cellules voisines, modifiant ainsi des frontières à des échelles inférieures à celle des arborisations. Nous pouvons donc résumer la situation en disant que, en général, aucun neurone n'est sélectionné isolément dans une carte, aucun neurone n'est connecté de façon réentrante à un unique neurone d'une autre carte ; qu'aucun neurone isolé ne présente les propriétés qu'il présente au sein d'un groupe. Toutes ces contraintes sont dues à la densité des connexions entre neurones, et elles font qu'il est extrêmement peu vraisemblable qu'un unique neurone puisse fonctionner comme unité de sélection.

Les trois principes de base de la TSGN en main, nous pouvons désormais examiner comment la capacité de catégoriser s'incarne dans le système nerveux. Pour le montrer, je vais prendre l'exemple de la catégorisation perceptive – la discrimination sélective d'un objet ou d'un événement parmi d'autres objets ou événements à des fins adaptatives. Je rappelle que cela ne relève pas d'une catégorisation classique, mais plutôt d'un échantillonnage disjonctif de propriétés. (On trouvera un passage détaillé sur les bases psychologiques de la catégorisation dans la section de la postface consacrée aux catégories.)

Pour tenter d'expliquer comment pourrait avoir lieu la catégorisation, nous pouvons utiliser les mécanismes de ce que j'ap-

pelle les « couples de classification » du cerveau. Il s'agit là d'une unité minimale, composée de deux cartes fonctionnellement distinctes comportant divers groupes neuronaux et connectées de façon réentrante (figure 9-4). Chacune des cartes reçoit, de façon *indépendante*, des signaux provenant soit d'autres cartes cérébrales, soit du monde extérieur (dans l'exemple présent, les signaux proviennent du monde extérieur). Pendant un certain intervalle de temps, l'émission de signaux réentrants relie fortement certaines combinaisons actives de groupes neuronaux appartenant à l'une des cartes à différentes combinaisons appartenant à l'autre carte. Cela se produit par renforcement ou affaiblissement des connexions synaptiques présentes au sein des groupes de chaque carte ainsi qu'au niveau de leurs connexions avec des fibres réentrantes. De cette manière, les fonctions et les activités de l'une des cartes sont reliées et corrélées à celles d'une autre carte. Ce processus a lieu en dépit du fait que chacune des cartes est en train de recevoir, indépendamment, des signaux provenant du monde extérieur – les uns, par exemple, pouvant être fournis par la vue et les autres par le toucher.

Si les cartes en question sont topographiquement connectées, elles vont pouvoir corréler des événements qui se déroulent à un endroit donné du monde extérieur sans qu'il soit besoin d'un superviseur d'ordre supérieur. (Le mot « topographique » qualifie une situation dans laquelle une couche de récepteurs sensoriels, recevant des signaux de l'extérieur, est reliée à la carte cérébrale correspondante de telle sorte que des emplacements voisins sur la couche sensorielle correspondent à des emplacements voisins sur la carte.) Cette connectivité ne se limite pas à un couple de cartes ou à un intervalle de temps donné : les interactions d'une multitude de cartes peuvent être coordonnées de la même manière.

Cette propriété est extrêmement importante. En effet, la sélection de groupes neuronaux qui s'opère à l'intérieur des cartes conduit à la production de nouvelles sortes de signaux, qui peuvent ensuite réentrer dans des cartes plus anciennes en même temps que des signaux provenant du monde extérieur. Par conséquent, la propriété de réentrée autorise ce que j'appelle une synthèse récursive : non seulement les événements sont corrélés topographiquement sur un ensemble de cartes sans l'aide d'un quelconque superviseur, mais des propriétés sélectives *nouvelles* émergent au cours du temps par réentrées successives et récursives à travers les cartes. Cette propriété a pu être simulée dans un modèle informatique du cortex

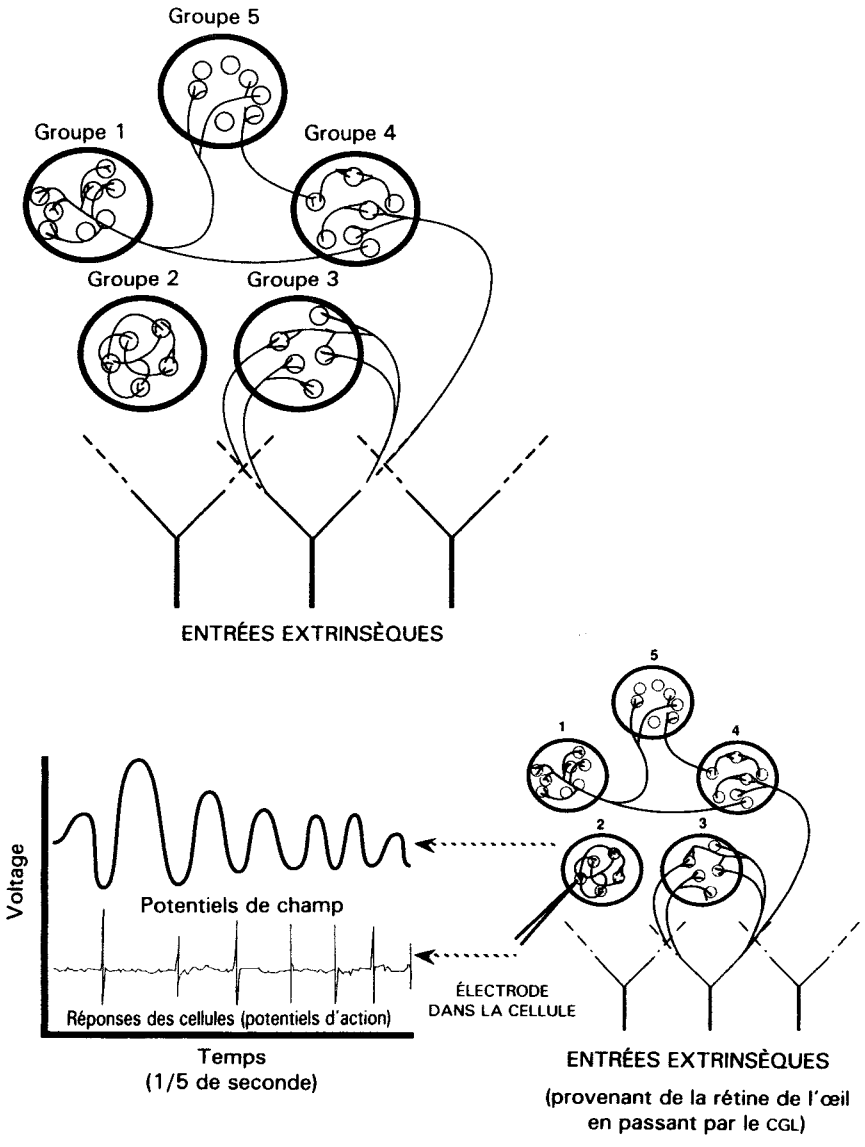


FIGURE 9-3 : *Les groupes neuronaux.* En haut *: les neurones se lient pour former des groupes (connexions intrinsèques) et les groupes se lient les uns aux autres (connexions extrinsèques). Chaque groupe permet de voir un aspect différent de la connectivité. Sur les groupes 1 et 5, on peut voir que chaque cellule établit des contacts avec des cellules de son propre groupe, ainsi qu'avec des cellules appartenant à d'autres groupes. Le groupe 2 permet de visualiser la densité de la connectivité intrinsèque des groupes. Avec le groupe 3, on voit que chaque groupe reçoit également des signaux provenant d'un ensemble de voies d'entrée extrinsèques qui se recouvrent partiellement, et qui peuvent être stimulées de façon sélective. (En général, ces voies d'entrée s'étendent sur des distances bien supérieures au diamètre cellulaire.) Le groupe 4 permet de voir que chaque cellule reçoit donc des entrées provenant de cellules de son propre groupe, de cellules appartenant à d'autres groupes et de sources extrinsèques. Les groupes diffèrent les uns des autres par leur taille (ils comportent entre environ 50 et 10 000 neurones) et par leur connectivité effective. Celle-ci est déterminée par la neuro-anatomie locale des aires dans lesquelles ils se trouvent.* En bas *: résultats confirmant l'existence de groupes neuronaux. Une électrode enregistre les réponses électriques (potentiels d'actions) d'un neurone visuel ainsi que celles de ses voisins au sein du groupe (potentiels de champ). Lorsqu'un stimulus visuel adéquat (une barre lumineuse se déplaçant vers le haut et vers la droite) est présent, les réponses du neurone et de ses voisins oscillent toutes à la même fréquence (40 Hertz, ou encore 40 cycles par seconde). Lorsqu'on interrompt le stimulus, les potentiels d'action et les potentiels de champ ne sont plus corrélés.*

cérébral, le modèle RCI [1], que j'ai déjà décrit en détail dans mon livre intitulé *The Remembered Present*. Ce modèle parvient à corréler, par réentrée, les activités de nombreuses cartes différentes et produit des réponses coordonnées à des images visuelles complexes.

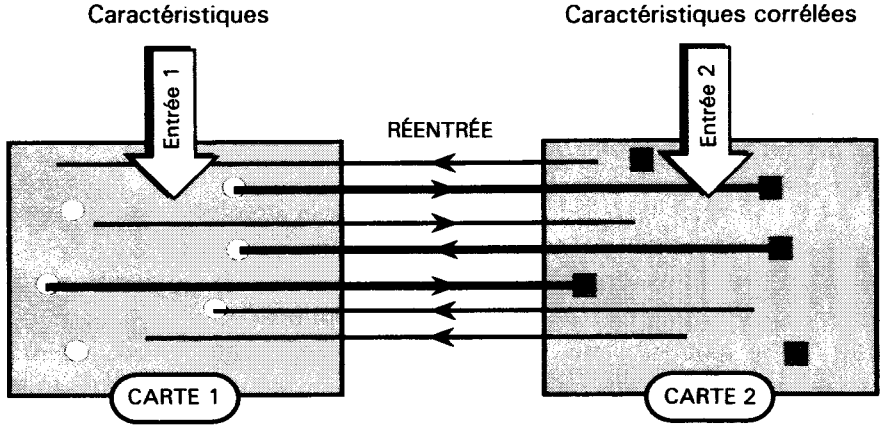


FIGURE 9-4

La réentrée. Deux cartes de groupes neuronaux reçoivent des entrées indépendantes (1 et 2). Chaque carte présente une ségrégation fonctionnelle ; autrement dit, la carte 1 réagit à des caractéristiques locales (à des angles détectés visuellement, par exemple) différentes de celles qui stimulent la carte 2 (qui réagit, par exemple, au mouvement d'ensemble d'un objet). Les deux cartes sont reliées par des fibres nerveuses qui transportent des signaux réentrants de l'une à l'autre. Ces fibres sont nombreuses et denses, et elles servent à « cartographier » les cartes l'une sur l'autre. Si, au cours d'un certain laps de temps, les groupes indiqués par des cercles dans la carte 1 sont connectés de façon réentrante aux groupes indiqués par des carrés sur la carte 2, ces connexions pourront être renforcées. À l'issue de l'émission de signaux réentrants, et par l'intermédiaire des modifications synaptiques, les réponses de la carte 1 se trouvent liées à celles de la carte 2, formant ce qu'on appelle un « couple de classification ». Les modifications synaptiques font également que les réponses à des stimulus actuels soient liées aux réponses antérieures.

Comment la réentrée peut-elle rendre compte de la catégorisation perceptive, fonction qui, selon la TSGN, est essentielle à toute tentative visant à relier la physiologie et la psychologie ? Énoncée en quelques mots, la réponse est la suivante : en couplant les sorties d'une multitude de cartes interconnectées de façon réentrante au comportement sensori-moteur de l'animal. Ce couplage est dû à une structure d'ordre supérieur, appelée cartographie globale. Une cartographie globale est une structure dynamique contenant de multiples cartes locales réentrantes (à la fois motrices et sensorielles)

1. *Reentrant cortical integration*, ou intégration corticale réentrante (N.d.T.).

capables d'interagir avec des parties non cartographiées du cerveau (figure 9-5). (Ces parties non cartographiées du cerveau correspondent notamment à des parties de certaines structures spécialisées connues sous le nom d'hippocampe, de ganglions de la base et du cervelet, dont je décrirai plus loin les fonctions.) Ici, je veux simplement mettre l'accent sur le fait que l'existence d'une cartographie globale permet à des événements sélectifs survenant dans les cartes *locales* de cette cartographie (du type de celles illustrées sur la figure 9-4) d'être reliés au comportement moteur de l'animal, à de nouveaux échantillonnages sensoriels du monde et à encore d'autres événements réentrants successifs.

Une telle cartographie globale garantit la création d'une boucle dynamique capable d'ajuster continuellement les gestes et la posture d'un animal aux échantillonnages indépendants de plusieurs sortes de signaux sensoriels. La sélection de groupes neuronaux au sein des cartes locales d'une cartographie globale aboutit alors à des réponses catégorielles particulières. Par conséquent, la catégorisation ne se produit pas selon un programme de type informatique dans une aire sensorielle qui ensuite exécute un programme pour fournir une sortie motrice particulière. Au contraire, c'est l'activité sensori-motrice sur l'ensemble de la cartographie qui *sélectionne* les groupes neuronaux donnant la sortie ou le comportement adéquats, ce qui permet d'aboutir à la catégorisation. Dans les systèmes de ce type, les décisions sont fondées sur la statistique des corrélations entre signaux. Remarquez à quel point cela tranche avec le fonctionnement d'un ordinateur : au lieu de dépendre du transport de messages codés contenant des instructions, ces modifications se produisent au sein d'un système sélectif.

Mais qu'est-ce qu'un comportement « adéquat », et comment la catégorisation perceptive se manifeste-t-elle ? Selon la TSGN, les catégorisations s'opèrent toujours par rapport à des critères de valeur internes, et c'est ce rapport qui définit leur adéquation. De tels critères de valeur ne déterminent pas des catégorisations spécifiques, mais ils imposent des contraintes aux domaines dans lesquels elles se produisent. D'après la théorie, chez les animaux d'une espèce donnée, les bases des systèmes de valeurs ont déjà été établies par sélection au cours de l'évolution. Et elles se manifestent dans les régions du cerveau qui participent à la régulation des fonctions corporelles : rythme cardiaque, respiration, comportement sexuel et alimentaire, fonctions endocrines, fonctions végétatives. La catégorisation se manifeste dans tous les comportements qui satisfont

de manière adéquate les exigences, sélectionnées par l'évolution, des systèmes physiologiques vitaux.

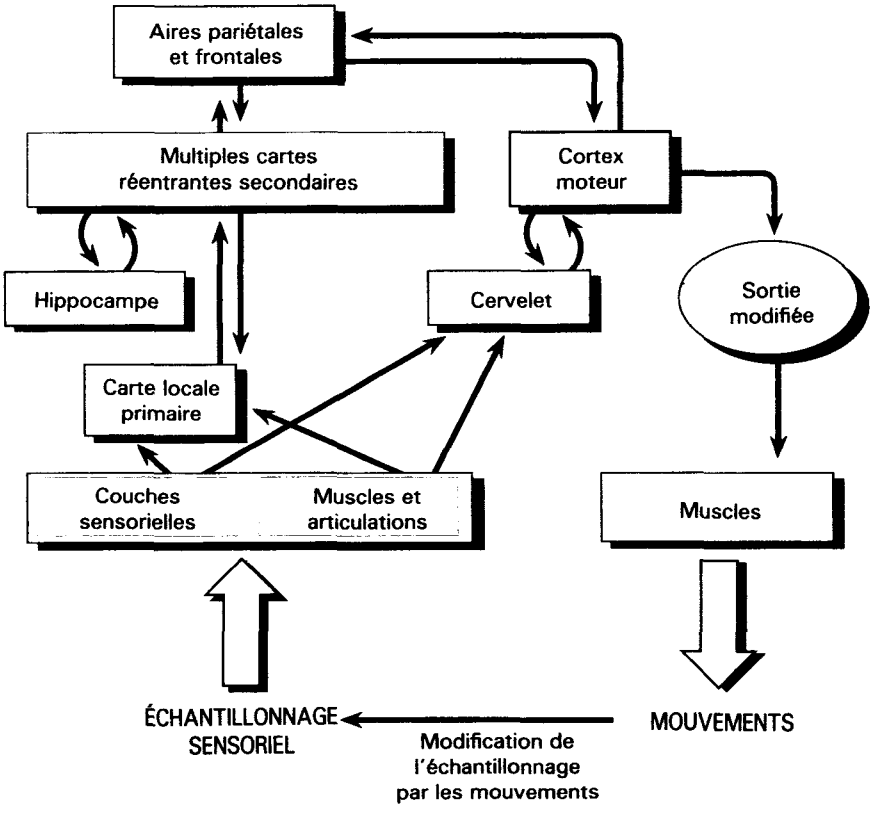


FIGURE 9-5 : *Une cartographie globale. Cette structure comporte de multiples cartes (du type de celles de la figure précédente). Les cartes sont également reliées à des régions du cerveau tel l'hippocampe et le cervelet. Remarquez que les signaux provenant du monde extérieur interviennent dans cette cartographie, et que les multiples sources de sorties aboutissent à des mouvements. Cela modifie la façon dont les signaux sensoriels sont captés. Ainsi, les cartographies globales sont des structures dynamiques, qui évoluent avec le temps et en fonction du comportement. Leurs cartes réentrantes locales, qui mettent en corrélation des caractéristiques et des mouvements, rendent possible la catégorisation perceptive. Des perturbations à différents niveaux peuvent entraîner la réorganisation ou l'effondrement d'une cartographie globale, ou encore son remplacement par une autre cartographie globale.*

Voici un exemple spécifique de catégorisation à laquelle des valeurs imposent des contraintes. Il pourra contribuer à relier toutes ces idées. Afin de montrer qu'il est possible d'effectuer des catégorisations perceptives d'après des valeurs dans les cartographies globales, mes collègues et moi-même avons simulé des automates complexes fondés sur la TSGN à l'aide de supercalculateurs (figure 9-6).

Dans des automates tels que Darwin III, on voit des valeurs opérer pour le système visuel, par exemple, dans les circuits qui favorisent l'éclairement de la partie centrale de l'œil. (Valeur = « la présence de lumière est mieux que l'absence de lumière » ; la présence de lumière et la stimulation au centre du champ visuel sont préférables à la présence de lumière et à la stimulation à la périphérie.) Chez Darwin III, l'action de ces circuits de valeur fait croître la probabilité que les synapses qui seront actives lorsque ces circuits entreront en jeu soient renforcées plutôt que les autres. Le résultat net est que, grâce à la sélection et à l'expérience, l'œil de l'automate parvient à suivre à la trace des signaux provenant d'objets éclairés.

Cela définit une forme de comportement « adéquat » comme étant un comportement acquis compatible avec des valeurs établies par l'évolution. Mais la catégorisation perceptive ne se produit que lorsque, après échantillonnage disjonctif des signaux appartenant à *plusieurs* modalités (vision, toucher, proprioception), Darwin III active une sortie par l'intermédiaire de ses cartes réentrantes. C'est le cas, par exemple, lorsque à l'issue d'une exploration effectuée par sa « main-bras » et son « œil », il « décide » que quelque chose est un objet, que cet objet a des rayures et qu'il est bosselé. Et comme Darwin III possède un système de valeurs d'ordre supérieur pour les sorties issues de ce type de décisions catégorielles, il active alors un circuit neuronal qui lui fait agiter le bras. Cette sortie reflète la catégorisation résultant des multiples événements sélectifs qui ont lieu sur les synapses de toutes ses cartes – connectées de façon réentrante – à l'issue d'une expérience. Ces événements sélectifs surviennent pendant les réactions de Darwin III et ne sont *pas* spécifiés par les valeurs. Comme on le voit sur la figure 9-6, Darwin III « prend des décisions » – qui se traduisent par des mouvements de son bras – à propos d'un vaste ensemble d'objets : il distingue ceux qui sont rayés et bosselés de ceux qui sont rayés ou bosselés, mais pas les deux choses à la fois. Tout comme un animal, il le fait d'une façon qui varie selon les individus, et non d'après des critères classiques prédéfinis. Autrement dit, il définit des catégories à partir de sa seule expérience, et non pas sur la base d'une programmation préalable. Darwin III est un modèle de cartographie globale qui *effectue des catégorisations sur des valeurs selon un mode que l'on pourrait qualifier d'incarné.* En général, une cartographie globale telle que celle de Darwin III représente la plus petite structure capable d'effectuer des catégorisations. Mais, bien entendu, le

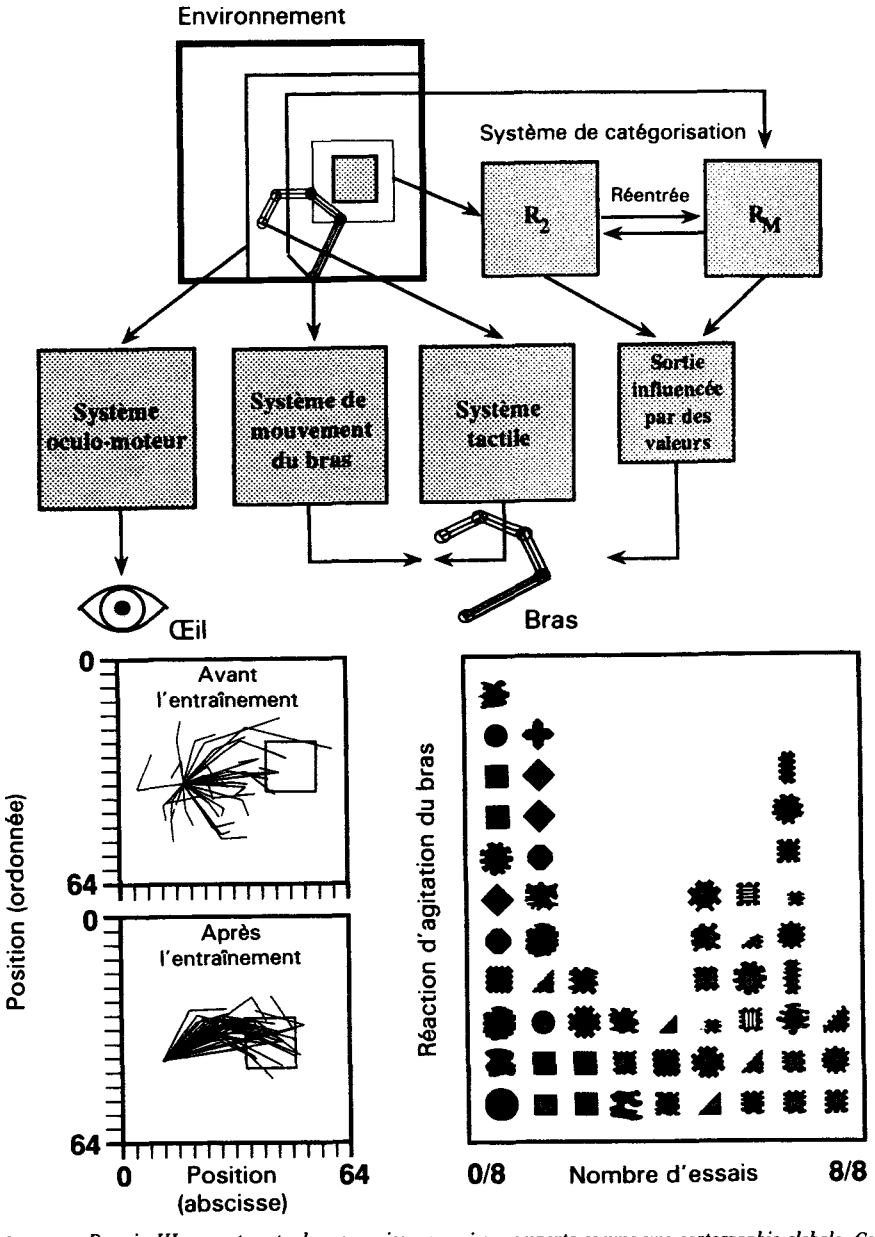


FIGURE 9-6 : *Darwin III, un automate de reconnaissance qui se comporte comme une cartographie globale. Cet automate a été simulé à l'aide d'un supercalculateur. Il possède un œil mobile unique, un bras à quatre articulations dont la dernière est sensible au toucher, et il est doté de kinesthésie (sens du mouvement des articulations), grâce aux signaux que lui envoient les neurones situés aux articulations au fil de leurs déplacements. Son système nerveux est structuré en plusieurs sous-systèmes, chacun étant chargé de différents aspects de son comportement* (en haut). *Il contient de nombreuses cartes du type de celles représentées sur les figures 9-4 et 9-5. Le phénotype « issu de l'évolution » – et notamment la neuro-anatomie – est* programmé *dans la simulation. Mais le comportement de la simulation n'est* pas *programmé (voir le chapitre 19). Une fois que l'automate a été confronté à des objets – qu'il les a « vus » se déplacer au hasard –, son œil devient capable de suivre n'importe quel objet. De même, l'automate tend le bras pour « toucher » les objets, et à chaque sélection de mouvements, il améliore sa capacité de les atteindre* (en bas à gauche). *Au cours des expériences représentées en bas à gauche, l'extrémité du bras se trouve toujours initialement au même endroit (le point de départ des tracés). On a enregistré graphiquement son mouvement en direction d'une région cible (le petit carré). Remarquez que, avant le début de l'entraînement, le bras se déplace dans de nombreuses directions. À l'issue d'un entraînement faisant intervenir une sélection* (l'ensemble de tracés du bas), *ses mouvements deviennent ciblés. Darwin III a été confronté à 55 objets différents, qui lui ont été présentés huit fois chacun et qu'il a dû classer par catégories. Les résultats présentés ici* (en bas à droite) *indiquent le nombre de fois (sur l'ensemble de huit présentations) où Darwin III a réagi de façon positive devant ces objets (c'est-à-dire en agitant le bras). Ces résultats montrent que Darwin III a divisé cet ensemble d'objets en deux classes.*

cerveau des vrais animaux possède la capacité de constituer un nombre bien plus élevé de cartographies de ce type.

Disons au passage que les mouvements du « bras » de Darwin III sont également le résultat d'une sélection et non une série d'instructions. Par exemple, en définissant la valeur comme un facteur qui fait croître la probabilité de voir se renforcer les connexions synaptiques dans les répertoires de mouvements qui sont actifs lorsque le bras de l'automate se déplace vers le centre du champ visuel, on obtient un système dans lequel des mouvements qui au départ n'étaient pas réglés finissent par céder la place à des mouvements capables d'atteindre un objet (figure 9-6).

Selon la TSGN, les forces motrices du comportement animal sont donc des ensembles particuliers de valeurs, sélectionnées au cours de l'évolution, qui aident le cerveau et le corps à maintenir les conditions nécessaires à la survie. Ces systèmes sont dits homéostatiques, et c'est le couplage du mouvement et de l'échantillonnage sensoriel, aboutissant à des comportements, qui modifie les niveaux des systèmes homéostatiques. Cependant, hormis les quelques types de comportement propres à chaque espèce qui ont été directement sélectionnés au cours de l'évolution, la plupart des catégorisations qui donnent lieu à des comportements capables de modifier les niveaux homéostatiques résultent de la sélection *somatique* de groupes neuronaux dans chaque animal. La catégorisation n'est pas une valeur – elle *se fonde* sur des valeurs. C'est un phénomène épigénétique du développement, et aucun circuit fondé sur des valeurs, aussi étendu soit-il, ne sera jamais susceptible d'aboutir à des catégorisations en l'absence d'une sélection de groupes neuronaux issue de l'expérience. Mais il est également vrai qu'en l'absence de valeurs préalablement définies, les systèmes somatiques de sélection ne convergeront pas vers des comportements précis. Cela a été démontré, par exemple, en coupant les circuits de valeurs de Darwin III : comme on pouvait s'y attendre, le comportement convergent illustré sur les figures ne survient pas dans ce cas.

Ce tour d'horizon aura été bien ardu. J'espère néanmoins avoir réussi à vous communiquer quelque peu la saveur et la logique interne de la TSGN. À présent, je voudrais dire quelques mots sur les données recueillies en faveur de la théorie depuis sa conception. Je n'ai pas l'intention d'être exhaustif, mais étant donné que la théorie a parfois été critiquée ou mal comprise, il me semble utile de clarifier certaines questions et de mentionner un certain nombre de résultats expérimentaux qui la corroborent.

Les deux concepts les plus vivement critiqués de la théorie sont les groupes neuronaux et la sélection elle-même. Horace Barlow et Francis Crick ont, chacun de leur côté, mis en cause l'idée de l'existence des groupes. La critique de Barlow repose sur l'idée que la formation de groupes neuronaux fait nécessairement appel à une dynamique malthusienne des populations. Thomas Malthus considérait que les populations croissent selon une progression géométrique alors que les réserves de nourriture ne le font que selon une progression arithmétique. Il influença Darwin, lequel raconte que ce fut en lisant Malthus qu'il comprit pourquoi la lutte pour la survie pouvait aboutir à la sélection naturelle. Mais le raisonnement de Barlow n'est pas aussi clair que celui de son éminent prédécesseur. Ni la sélection naturelle qui s'opère au cours de l'évolution ni la sélection des groupes neuronaux n'ont obligatoirement besoin d'une population en croissance. Ce qu'il leur faut, c'est une reproduction *différentielle* (dans le cas de l'évolution) et une amplification *différentielle* (dans le cas de la sélection des groupes neuronaux par modification de la force des liaisons synaptiques). En fait, l'erreur de Barlow est double, car il se fonde sur un mécanisme erroné de modification synaptique pour conclure qu'au cours de la sélection, soit les groupes de neurones incorporent un nombre croissant de cellules soit ils perdent leur sélectivité. Or des modèles explicites ont permis de montrer qu'aucune de ces conséquences n'est vraie si l'on choisit des règles de modification synaptique qui ressemblent davantage à celles que l'on trouve dans les expériences réelles.

Quant à Crick, il affirme qu'il existe peu de données expérimentales en faveur de l'existence des groupes neuronaux. Il prétend également qu'il n'est pas nécessaire de faire appel à la sélection des groupes neuronaux pour dire qu'il existe des cartographies globales. Enfin, il affirme qu'il n'a pas réussi à établir une comparaison valable entre la théorie de la sélection naturelle et ce qui se passe dans le cerveau en cours de développement.

Cependant, des résultats expérimentaux contraires à ces affirmations ont été obtenus depuis la conception de la TSGN. Ils constituent des preuves directes de l'existence de groupes neuronaux et des fonctions de réentrée. L'un d'eux concerne le comportement des neurones dans les « colonnes d'orientation » du cortex visuel primaire (voir les figures 9-2 et 9-3). On sait que lorsque l'on présente des barres lumineuses mobiles au système visuel d'un animal, des neurones particuliers de ces colonnes répondent, en déchargeant des influx nerveux, aux barres ayant une orientation

particulière. Différents neurones sont « syntonisés » sur différentes orientations ; ils réagissent spécifiquement à des barres ayant une certaine orientation et se déplaçant dans une certaine direction. Wolf Singer et ses collègues, et aussi Reinhardt Eckhorn et ses collègues, ont montré que, dans cette partie du cortex visuel, le fait qu'une cellule prise isolément décharge ou non est très fortement corrélé avec la réponse des cellules qui sont ses voisines sur la même colonne. L'activité de ces cellules voisines a été mesurée en enregistrant des potentiels de champ, qui représentent l'activité cumulée d'un grand nombre de cellules prises dans une petite région. Ces expériences ont montré que la présentation d'une barre ayant l'orientation adéquate suscitait des décharges simultanées chez un *groupe* de cellules situées dans la colonne répondant à cette orientation, avec une composante oscillatoire prédominante à 40 Hertz (40 cycles par seconde ; voir la figure 9-3). Lorsqu'on enlevait le stimulus, la réponse oscillatoire cohérente du groupe de cellules disparaissait. De manière tout aussi surprenante, lorsqu'on présentait une barre et qu'on enregistrait les potentiels de champ dans deux cartes visuelles *distinctes*, V1 et V2 (voir la figure 9-2), les groupes de neurones des deux cartes oscillaient en phase, à environ 40 Hertz. Autrement dit, malgré la distance qui les séparait, des groupes de neurones appartenant à des cartes distinctes oscillaient à la même fréquence et en phase sur présentation de la barre lumineuse. Or on sait que les neurones de ces groupes éloignés sont reliés par des fibres réentrantes. Ces résultats semblent donc confirmer l'idée, avancée dans la première version de la TSGN, de l'existence d'un échange réentrant de signaux en phase. Mes collègues et moi-même avons modélisé ces résultats à l'aide d'un ordinateur. Et nous avons constaté qu'il suffisait de couper une seule branche à la voie réentrante entre les deux aires pour que les oscillations des groupes neuronaux simulés correspondants ne soient plus en phase et deviennent incohérentes.

Ces résultats corroborent : (1) la notion de groupe – des neurones plus ou moins fortement couplés par des connexions synaptiques, interagissant conjointement, déchargeant en même temps des influx nerveux et répondant unitairement à la sélection opérée par des stimulus particuliers ; et (2) l'idée de mise en corrélation par réentrée des événements sélectifs survenant dans différentes cartes. Il a été démontré expérimentalement que ces deux idées sont valides dans le cas de répertoires secondaires tels que ceux du système visuel.

Nous avons également réussi à modéliser les modifications plastiques qui s'opèrent au niveau des frontières des cartes d'un autre répertoire secondaire, situé dans la partie du cortex associée au toucher. Ces modifications furent découvertes au cours d'expériences réalisées par Michael Merzenich et ses collègues (voir le chapitre 3, en particulier le bas de la figure 3-5). L'idée que des groupes neuronaux entrent en *compétition* pour parvenir à interagir avec des neurones voisins en renforçant leurs connexions synaptiques a parfaitement permis d'expliquer cette plasticité. Ainsi, le fait de modifier les suites de légers stimulus tactiles ou encore de couper les nerfs responsables de la sensibilité tactile des doigts conduit à des modifications rapides des frontières des cartes du cortex cérébral somato-sensoriel assurant ces fonctions. Ces résultats sont entièrement compatibles avec la notion darwinienne de sélection parmi des groupes en concurrence au sein d'une carte.

Il semble donc que les critiques dirigées contre la notion de groupe neuronal ne tiennent pas debout. De même, les critiques à propos de la formation des répertoires primaires révèlent une profonde incompréhension. En particulier, l'interprétation de Dale Purves, selon laquelle la théorie est « régressive » parce qu'elle suggère que la sélection des cellules nerveuses n'opère que par élimination au cours du développement, constitue tout simplement une déformation des choses. Dans la description de la formation des répertoires primaires présentée dans *Neural Darwinism*, je dis explicitement qu'une sélection éliminatoire ne suffirait pas. En effet, bien que ce type de sélection ait indubitablement lieu durant la formation du système nerveux, elle ne constitue qu'un mécanisme sélectif parmi tant d'autres. La formation de nouvelles voies anatomiques par expression de nouvelles molécules d'adhérence et celle de boucles de transmission de signaux par établissement de synapses en fonction de l'activité sont des mécanismes sélectifs au moins tout aussi importants.

Le point de vue de Crick (voir la citation qui figure au début de ce chapitre), qui pense que la théorie ne devrait pas s'appeler darwinisme neuronal, mais plutôt edelmanisme neuronal parce qu'elle n'a rien à voir avec le travail de Darwin, est à la fois dérisoire et flatteur. Il est également déplacé. Comme l'a fait remarquer Richard Michod, et comme je l'ai dit moi-même, cette théorie présente des parallèles très nets avec les concepts darwiniens. Autrement dit, elle fait appel au raisonnement sur des populations d'une façon tout à fait rigoureuse. Par exemple, chaque groupe neuronal

présente une probabilité de répondre à une entrée, et cette probabilité est liée à des caractéristiques structurelles qui sont sujettes à des variations, ce qui correspond bien à l'idée d'adaptation. Plus précisément, la connectivité particulière de chaque groupe influe directement sur cette probabilité. De plus, il existe une relation entre la notion d'héritabilité et la sélection des groupes neuronaux. En effet, dans un système sélectif, il doit exister une corrélation supérieure au bruit de fond entre les entités parents et leur progéniture. Dans le cas de l'évolution, ce phénomène est garanti par l'hérédité, et dans la TSGN par les modifications synaptiques. Ainsi, en moyenne, les groupes neuronaux qui, au départ, répondent à un stimulus, ont une plus grande probabilité de répondre ultérieurement à un stimulus semblable, bien que cette probabilité soit modulée par les systèmes de valeurs. Dans le cas de l'évolution, les différences d'adaptation des divers organismes à l'environnement aboutissent à des différences dans les processus reproductifs, qui à leur tour conduisent à des modifications de la fréquence des gènes dans la population. Dans le cas de la sélection des groupes neuronaux, les différences au niveau de la connectivité, de la structure synaptique et de la morphologie des neurones du répertoire primaire conduisent, après confrontation à différentes suites corrélées de signaux provenant du milieu extérieur, à des différences au niveau de la probabilité de voir ces neurones réagir en tant que groupes. Cela reflète des modifications dans la structure de leurs forces synaptiques. Dans un cas, on parle de reproduction différentielle, et dans l'autre d'amplification différentielle. L'erreur de Crick est en fait probablement due à l'idée fausse selon laquelle il est impossible que les éléments d'un répertoire varient à la suite d'une sélection et que, par conséquent, ce sont nécessairement des entités absolument figées. Or cela n'est vrai ni pour la sélection naturelle, ni pour la sélection des groupes neuronaux.

Ce qu'il est crucial de reconnaître c'est que, tandis que les sciences de la reconnaissance (l'évolution, l'immunologie, la neurobiologie) sont fondées sur des *principes* communs, leurs *mécanismes* sont évidemment différents. Mais ce que tout cela a d'extrêmement surprenant, c'est le fait que la *sélection naturelle* ait produit, au cours de l'évolution des espèces, deux systèmes de *sélection somatique* complètement différents capables d'effectuer une reconnaissance. Si nous acceptons ces idées, nous voyons qu'une petite boucle comprenant les diverses étapes de la sélection des groupes neuronaux conduit à des comportements phénotypiques diversifiés chez les

différents individus d'une espèce. Ces divers comportements servent de base à la sélection naturelle en cours dans la grande boucle de l'évolution. Les deux systèmes sélectifs, le somatique et l'évolutif, interagissent.

En matière de recherche scientifique, tout doit être sujet à des critiques impitoyables. Ce qui est néanmoins curieux dans les critiques formulées contre la TSGN, c'est le niveau auquel elles se placent. On aurait pu s'attendre à ce que la plupart d'entre elles s'attaquent à la tentative pour établir un pont entre la psychologie et la physiologie – c'est-à-dire aux mécanismes proposés pour rendre compte de la catégorisation perceptive et de la mémoire. Ces mécanismes, comme tous les mécanismes de formation de concepts envisagés – que j'aborderai plus loin – sont vraiment au cœur de la théorie et n'ont pas encore été mis à l'épreuve. Au point où nous en sommes aujourd'hui, cependant, ni les résultats expérimentaux sur lesquels se fonde la TSGN, ni les propres propositions de la théorie n'ont été supplantés. Au contraire, certaines des prédictions issues de la théorie ont même déjà été confirmées. En fait, il serait immensément utile que des chercheurs prouvent que les faits que j'ai présentés sont faux, ou encore qu'une théorie alternative soit bientôt formulée à partir de ces mêmes faits. Nous aurions alors un réel espoir de voir apparaître des critiques plus constructives et des progrès qui pourraient affiner davantage notre vision de la fonction cérébrale.

Jusqu'ici, cela n'a pas été le cas. Je vais par conséquent aborder maintenant les parties de la théorie qui traitent de la façon de réduire l'abîme qui sépare la physiologie de la psychologie. Toute description biologique de la conscience passe nécessairement par là.

Chapitre 10

Mémoire et concepts : un pont vers la conscience

« Concept » est un concept vague.
Ludwig Wittgenstein

Qu'est-ce qu'une idée ? C'est une image qui se peint dans mon cerveau.
Voltaire

Le moment est venu de faire le point et de regarder en avant. Nous avons passé un certain temps à essayer de voir comment s'incarnent les fonctions mentales, à tenter de relier la psychologie à la physiologie. Nous avons dit que la sélection naturelle a donné naissance aux systèmes somatiques sélectifs – au système immunitaire et au cerveau – que la fonction cérébrale se fonde principalement sur la morphologie et que la neuro-anatomie adéquate se développe selon des principes topobiologiques. De fait, le cerveau est un système topobiologique par excellence, puisqu'il se compose de cartes et de systèmes de cartographie dans lesquels la position est essentielle au bon fonctionnement.

Deux constatations apparemment sans rapport entre elles nous ont obligés à réexaminer la façon dont le cerveau pourrait fonctionner en tant que système de reconnaissance. La première est l'immense diversité et l'individualité de la structure cérébrale. La seconde est que le monde, bien que soumis aux contraintes imposées par les lois de la physique, est un lieu dépourvu d'étiquettes. Pour tenter d'expliquer la façon dont un cerveau ainsi constitué pourrait parvenir à classer un tel monde par catégories, nous avons formulé la TSGN. Les principes de base de cette théorie – la sélection au cours du développement, la sélection par l'expérience, et la réentrée – sont considérés comme étant les bases essentielles au développement des fonctions psychologiques. Cela ne signifie pas, cependant, que de nouveaux agencements morphologiques ne soient pas nécessaires aux fonctions psychologiques *naissantes*, mais simplement que la TSGN part de l'idée qu'il n'est besoin d'adjoindre aucun autre principe majeur pour qu'apparaissent de nouvelles fonctions

au cours de l'évolution. Je me propose de montrer ici que l'action de la sélection somatique sur les cartographies globales, alliée à l'adjonction de nouveaux types de cartographies au cours de l'évolution, constitue un puissant moyen d'acquisition de nouvelles fonctions, telles que des mémoires spécialisées et des aptitudes conceptuelles.

Avant de nous intéresser à la mémoire et aux concepts proprement dits, je pense que nous devons nous demander comment nous concevons les relations existant entre les diverses « fonctions cérébrales supérieures ». Quelles sont les fonctions psychologiques dont le point de vue sélectionniste devrait rendre compte ? Et comment doivent-elles à leur tour rendre compte de la conscience et de l'intentionnalité ?

La triade fondamentale des fonctions cérébrales supérieures est formée par la catégorisation perceptive, la mémoire et l'apprentissage. (Bien que ces fonctions soient souvent traitées séparément pour des raisons de commodité, nous devons garder présent à l'esprit qu'en fait, elles représentent des aspects indissociables d'une capacité mentale globale.) Nous avons déjà vu comment les couples de classification et les cartographies globales effectuent des catégorisations perceptives. Ces catégorisations sont généralement nécessaires à la mémoire, qui n'est, après tout, qu'une affaire de catégorisations plus anciennes. Le fonctionnement à la fois de la catégorisation perceptive et de la mémoire peut être testé en analysant les comportements. Par ailleurs, nous verrons au cours de ce chapitre que bien qu'elles soient fondées sur des modifications des forces synaptiques, toutes les mémoires sont des propriétés des systèmes dynamiques dont les caractéristiques dépendent des structures neuronales concrètes dans lesquelles elles se manifestent. Toutefois, pour que les besoins adaptatifs d'un animal confronté à des juxtapositions imprévues d'événements qui conditionnent sa survie soient satisfaits, il faut également que l'apprentissage influe sur le comportement. Ainsi, les trois fonctions fondamentales – la catégorisation, la mémoire et l'apprentissage – sont intimement liées : la dernière dépend des deux premières.

Pourtant, alors que la catégorisation perceptive et la mémoire sont nécessaires à l'apprentissage, elles ne sont pas suffisantes. Encore faut-il qu'elles soient liées à des systèmes de valeurs par l'intermédiaire de parties du cerveau différentes de celles qui effectuent les catégorisations. La condition suffisante à l'adaptation est fournie par la liaison des cartographies globales à l'activité de ce qu'on

appelle les centres hédonistes et le système limbique du cerveau, de façon à satisfaire les besoins homéostatiques, alimentaires et sexuels reflétant les valeurs établies au cours de l'évolution. Les diverses structures cérébrales chargées de valeurs – notamment l'hypothalamus, divers noyaux du mésencéphale et autres – sont apparues au cours de l'évolution des espèces en réponse à des exigences éthologiques, et certains de leurs circuits sont spécifiques de l'espèce considérée. Cela pour une raison évidente : l'activité sexuelle et le comportement des oiseaux sont très différents de ceux des baleines.

Dans toutes les espèces, l'apprentissage résulte du fonctionnement des liaisons neuronales existant entre les cartographies globales et les centres de valeurs cités ci-dessus. Il sert à relier les catégorisations à des comportements qui ont une valeur adaptative dans des conditions d'attente. Les systèmes physiologiques, tout comme certains dispositifs de contrôle, ont des points fixes (pensez à un thermostat). Dans ce contexte, une condition d'attente est tout simplement une condition dans laquelle les points fixes des structures physiologiques qui font partie du système hédoniste n'ont pas encore été atteints. L'apprentissage est achevé lorsque le comportement conduit à des modifications synaptiques dans les cartographies globales qui permettent d'atteindre les points fixes.

Nous voyons à présent pourquoi le fonctionnement de la mémoire, dans la mesure où il relie la catégorisation perceptive à l'apprentissage, souligne la valeur adaptative de la sélection des groupes neuronaux. L'accroissement de la taille des répertoires primaires ou de la connectivité réentrante entre répertoires, ou encore le renforcement des moyens de modification des liaisons synaptiques par adjonction de nouveaux mécanismes chimiques au cours de l'évolution, font augmenter le nombre de réponses catégorielles susceptibles de renforcer l'apprentissage. Et comme l'apprentissage est adaptatif, ce raisonnement permet de dire que plus les groupes neuronaux sont nombreux et diversifiés, plus ils devraient être adaptatifs. Toutefois, quel que soit le degré d'apprentissage, les comportements sont soumis aux contraintes imposées par des facteurs éthologiques – dont les systèmes de valeurs et les exigences homéostatiques, sélectionnés au cours de l'évolution d'une espèce, sont parmi les plus importants.

La mémoire se trouve au centre de tous ces événements, et je vais consacrer une bonne partie de la suite de ce chapitre à l'analyse de ses mécanismes et de ses besoins. Cependant, pour importante que puisse être la triade de base constituée par la perception, la

mémoire et l'apprentissage, leur fonctionnement conjoint ne suffit pas à engendrer les types de capacités qui permettent de relier entre elles les différentes catégorisations perceptives pour aboutir à des propriétés *relationnelles* générales. Ces propriétés sont issues de l'acquisition de compétences conceptuelles, c'est-à-dire de la capacité d'établir des catégories en termes de relations générales ou abstraites. Par conséquent, il faudra également que j'aborde la question des concepts. Mon but est de montrer comment, une fois assises les bases physiologiques de notre triade centrale et celles des aptitudes conceptuelles, nous pouvons rendre compte de l'émergence de la conscience en ne faisant appel qu'aux principes déjà présents dans la TSGN.

LA MÉMOIRE

Commençons par la mémoire. L'une des difficultés auxquelles on est confronté lorsqu'on veut aborder ce sujet provient du grand nombre de types différents de mémoire qui ont été décrits : comme ils sont si nombreux à être intimement liés aux capacités linguistiques, on a beaucoup de mal à en discerner les éléments fondamentaux. Il en résulte que l'on assimile souvent, à tort, l'une des bases physiologiques de la mémoire – les modifications synaptiques – à la mémoire elle-même.

Pour clarifier les idées à ce propos, nous allons convenir que, quelle que soit la forme qu'elle adopte, la mémoire sera toujours définie comme la capacité de répéter une performance. Le type de performance dont il s'agit dépend de la structure du système dans lequel la mémoire se manifeste, car la mémoire est une propriété systémique. En tant que telle, dans le système nerveux, c'est une propriété dynamique des populations de groupes neuronaux. Dans les ordinateurs, elle dépend de la spécification et du stockage de bits d'information codée. Et, selon la TSGN, tel n'est *pas* le cas dans le système nerveux. Ainsi par exemple, alors que le comportement de Darwin III donne des signes évidents d'une certaine forme de mémoire à la suite de ses rencontres avec divers objets (voir le chapitre précédent, en particulier la figure 9-6), sa mémoire ne peut être décrite en termes de configurations statiques de bits. La mémoire dont font preuve les cartographies globales telles que Darwin III ne provient pas du stockage d'attributs codés, fixes, appelés

et rassemblés de manière réplicative, comme c'est le cas pour un ordinateur.

La TSGN postule que la mémoire consiste plutôt en un renforcement spécifique d'une capacité de catégorisation préalablement établie (figure 10-1). Ce type de mémoire est une propriété collective issue des modifications dynamiques survenant continuellement dans les populations synaptiques des cartographies globales – et qui sont du même type que celles qui ont permis à des catégorisations d'avoir lieu au départ. La modification des forces synaptiques des groupes dans une cartographie globale constitue la base biochimique de la mémoire.

Dans un tel système, la remémoration n'est pas stéréotypée. Sous l'influence de contextes qui se modifient constamment, elle change à mesure que la structure et la dynamique des populations neuronales impliquées dans les catégorisations originelles changent elles aussi. La remémoration fait appel à l'activation de certaines portions – mais pas nécessairement à la totalité – des cartographies globales établies auparavant. Elle peut ainsi donner lieu à des catégorisations semblables à des catégorisations précédentes, mais, à des instants différents, les éléments contribuant à cette réponse seront eux-mêmes différents, et en général ils seront susceptibles d'être modifiés par les comportements survenus entre-temps.

Puisque les catégories perceptives ne sont pas immuables et qu'elles se modifient sous l'effet des comportements de l'animal, la mémoire, vue sous cet angle, résulte d'un processus de continuelle *recatégorisation.* Par nature, elle fait intervenir des procédures, une activité motrice continuelle et des essais répétés dans divers contextes. Et, étant donné que de nouvelles associations peuvent apparaître dans ce contexte, que les entrées et les stimulus se modifient, et que différentes combinaisons de groupes neuronaux sont susceptibles de produire des sorties semblables, une réponse catégorielle donnée de la mémoire peut être atteinte de plusieurs façons. Ainsi, contrairement aux mémoires électroniques, la mémoire cérébrale est imprécise, mais elle possède, en revanche, de grandes capacités de généralisation.

Les propriétés d'association, d'imprécision et de généralisation découlent toutes du fait que la catégorisation perceptive, qui est l'une des bases premières de la mémoire, est de nature probabiliste. Il n'est donc pas surprenant que différents individus puissent avoir des souvenirs aussi différents et qu'ils les utilisent de façon si diverse.

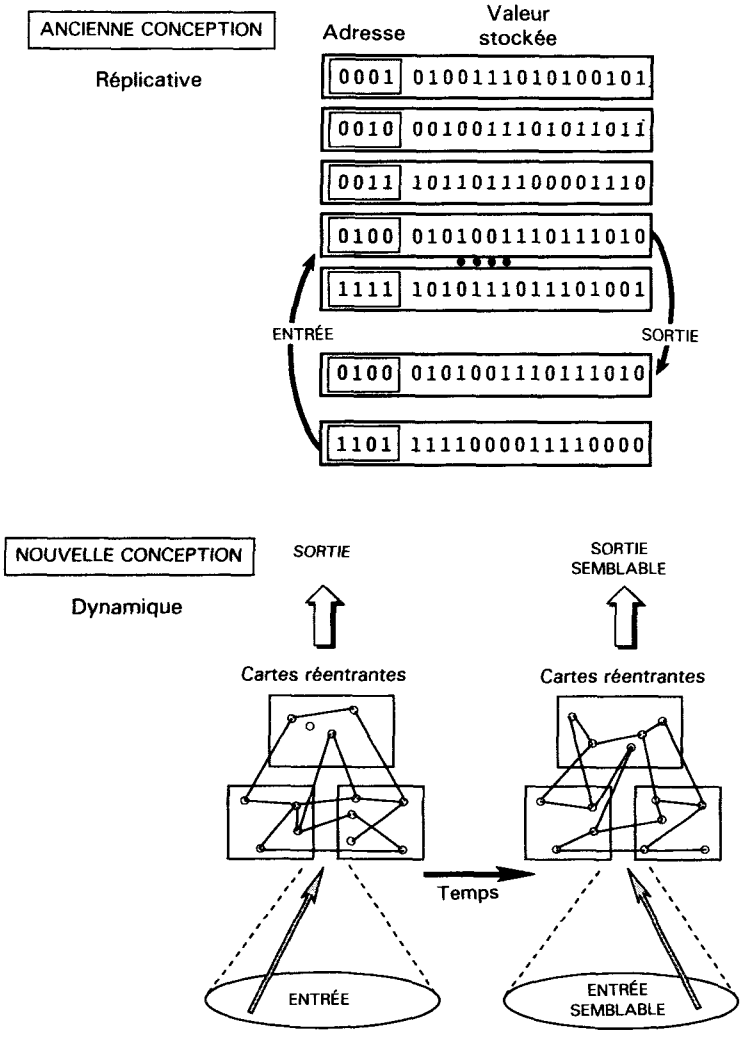


FIGURE 10-1

Deux conceptions de la mémoire. En haut : *exemple de mémoire vue sous l'angle du stockage d'informations précisément codées (mémoire réplicative). Je l'appelle réplicative parce que le rappel de cette mémoire doit reproduire la même forme codée sans erreur et donc la « répliquer » fidèlement, à l'instar d'un ordinateur. La simple modification d'un bit d'information, où que ce soit, constitue une erreur.* En bas *: exemple de mémoire dynamique dans une cartographie globale du type de celle de Darwin III (figure 9-6), à l'issue de catégorisations fondées sur des valeurs. De nombreux objets, s'ils sont catégorisés de la même manière, peuvent donner des sorties semblables, et des erreurs peuvent donc être commises. Cette mémoire* constitue une propriété de l'ensemble du système, *bien que son mécanisme de base soit la modification des efficacités synaptiques, comme l'indiquent les modifications intervenues dans les segments qui joignent les groupes neuronaux (petits cercles) au sein des cartes.*

En décrivant ces propriétés, je me suis rappelé à quel point le fonctionnement de la mémoire et les performances des individus de talent peuvent être différents. Voici une anecdote que l'on raconte à propos de Fritz Kreisler, le grand violoniste, et de Sergheï Rachmaninov, le grand pianiste. Elle illustre bien ce que je veux dire. En 1930, ces deux hommes se rencontrèrent à Berlin pour enregistrer ensemble la sonate en do mineur de Grieg. Rachmaninov, qui était très méticuleux dans son travail, voulut commencer aussitôt les répétitions. Mais Kreisler, qui n'avait pas l'habitude de beaucoup répéter, ne se montra pas aussi empressé et partit se promener en ville. Le lendemain matin, du fait de l'insistance de Kreisler et malgré les réticences de Rachmaninov, l'enregistrement eut lieu, et tout se passa bien. (Je crois que l'enregistrement est encore disponible, et il est sensationnel.) Quoi qu'il en soit, Rachmaninov ne fut pas content.

L'histoire veut qu'un peu plus tard cette même année-là, les deux hommes aient donné ensemble un concert, à New York, au programme duquel figurait la même sonate. Au milieu d'un des mouvements, Kreisler eut un trou de mémoire. Comme à son habitude, il inventa simplement un certain nombre de mesures, probablement dans l'espoir de reprendre le fil de l'œuvre un peu plus loin. Mais comme, passé une minute environ, il n'y était toujours pas parvenu, il se pencha vers Rachmaninov tout en continuant à jouer, et lui demanda : « Sergheï, où est-ce que nous sommes ? » Rachmaninov leva les yeux du clavier et lui répondit : « À Carnegie Hall. »

Si l'on considère que la mémoire est une forme de recatégorisation, il est évident que l'on ne peut comprendre ses mécanismes qu'en considérant l'ensemble du système dans lequel elle opère. (On pourra trouver un exemple en se reportant à Darwin III au chapitre précédent.) L'une des caractéristiques dynamiques du système de cartographies globales du cerveau est sa capacité d'ordonner dans le temps des changements successifs. De fait, la mémoire ne serait d'aucune utilité si elle ne pouvait pas, d'une façon ou d'une autre, tenir compte de l'enchaînement des événements dans le temps – aussi bien en ce qui concerne les événements sensoriels que les suites de comportements moteurs.

Pour voir comment tout cela fonctionne, il nous faudrait plonger dans un océan de détails techniques. Néanmoins, il est important de comprendre quels sont les moyens mis en œuvre par le cortex cérébral et ses appendices pour satisfaire les exigences temporelles et spatiales de la mémoire. Je rappelle que le cortex est une structure

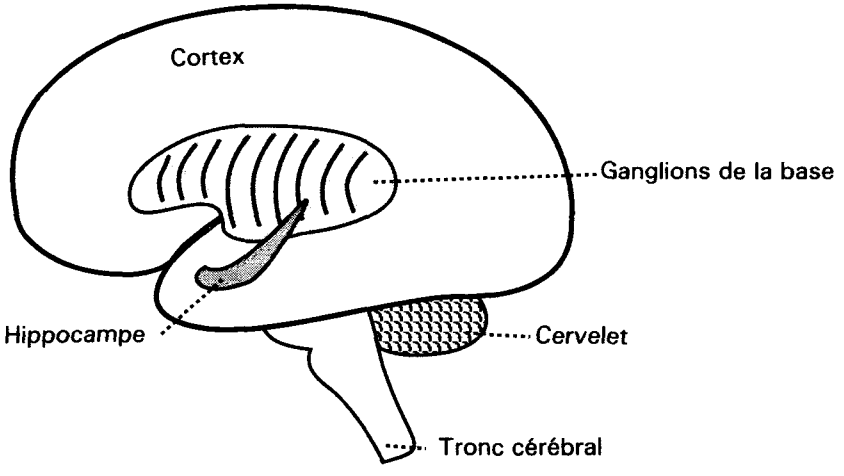
constituée de six couches interconnectées, contenant quelque dix milliards de neurones et environ un million de milliards de connexions. Outre le fait qu'il est structuré en cartes aux fonctions distinctes, connectées de façon réentrante et desservant la totalité des modalités sensorielles et des réponses motrices, il est relié à trois structures, que j'appelle les organes de la succession – parce qu'ils interviennent, bien entendu, dans l'organisation des sorties du cerveau.

Chacune de ces structures – le cervelet, l'hippocampe et les ganglions de la base – traite d'un aspect différent et d'une échelle différente de cette organisation (figure 10-2). Le cervelet est une structure remarquable qui entoure le tronc cérébral supérieur. Elle consiste en un ensemble caractéristique de circuits neuronaux à la structure plutôt stéréotypée, et il reçoit principalement deux types d'entrées, provenant du cortex cérébral et de la moelle épinière. Diverses études incitent à penser que, bien que le cervelet ne soit pas absolument indispensable au déclenchement des mouvements, il joue un rôle extrêmement spécifique dans la synchronisation et le déroulement sans saccades des suites de mouvements. Avec le cortex cérébral, il est à la base de la production et de la catégorisation de gestes sans saccades. En l'absence de certaines portions du cervelet, les suites de mouvements qui normalement s'enchaînent parfaitement deviennent saccadées et perdent leur coordination.

Mais qu'est-ce que tout cela a à voir avec la mémoire ? Rappelez-vous : la catégorisation dépend du bon enchaînement des gestes et des postures autant que des couches sensorielles. Et le cervelet et le cortex moteur sont tous deux soumis aux modifications synaptiques qui aboutissent au parfait fondu enchaîné moteur sous-jacent à la fois à la catégorisation et à la recatégorisation.

L'exécution de séquences d'événements moteurs plus longues, appelées programmes moteurs, dépend d'un autre ensemble d'appendices corticaux : les ganglions de la base. Il s'agit d'un grand ensemble de structures complexes enfouies au centre du cerveau. Ces structures sont non seulement reliées au cortex cérébral par l'intermédiaire d'une série de circuits parallèles qui participent aux mouvements oculaires et corporels, mais aussi aux portions frontales du cortex, dont la fonction est liée à la planification des comportements et aux émotions. Les ganglions de la base interviennent dans la planification des mouvements et par conséquent dans le choix des types et des enchaînements de sorties motrices. Non

seulement ils contribuent à la régulation des mouvements dans les programmes moteurs en couplant les réponses sensorielles et motrices, mais ils aident également à orienter ce qui doit être fait selon un plan moteur donné. Vous remarquerez que, contrairement au cervelet, qui enchaîne et coordonne des gestes plus immédiats, cet appendice fonctionne sur des échelles de temps plus grandes et contribue à corréler des séquences entières de gestes au sein d'un plan.



APPENDICES CORTICAUX

FIGURE 10-2

Les appendices corticaux – les organes de la succession. Le cerveau contient des structures telles que le cervelet, les ganglions de la base et l'hippocampe, qui interviennent dans la synchronisation, la succession des mouvements et l'établissement de la mémoire. Ces organes entretiennent des relations étroites avec le cortex cérébral pendant que celui-ci effectue des catégorisations et établit des corrélations du type de celles assurées par les cartographies globales (voir la figure 9-5). Ce diagramme sert simplement à aider le lecteur à situer ces « organes de la succession » au sein du cerveau.

Le rôle central des ganglions de la base dans les mouvements et les plans moteurs est mis en évidence dans certaines maladies. Dans la maladie de Parkinson, par exemple, la destruction d'un ensemble particulier de neurones des ganglions de la base (la substance noire), qui fabriquent un neuromédiateur appelé dopamine, conduit à des difficultés dans le déclenchement des mouvements, à des tremblements et à des perturbations de la marche. Par ailleurs, les ganglions de la base sont intimement liés aux centres hédonistes du cerveau

et, comme nous le verrons plus loin, ils jouent très probablement un rôle dans l'attention.

Un troisième appendice cortical très important, l'hippocampe, est lui aussi intimement lié aux centres hédonistes situés dans le mésencéphale et dans l'hypothalamus. Sa principale caractéristique réside dans le fait qu'il joue un rôle important dans la liaison entre la mémoire à court terme et la mémoire à long terme. L'hippocampe siège sur la face interne du cortex temporal, en bordure du tronc cérébral. C'est une structure en forme de boudin dont la section longitudinale ressemble à un C (c'est sa ressemblance avec les hippocampes marins qui lui a valu son nom).

L'hippocampe a ceci d'extraordinaire qu'il reçoit des entrées provenant de pratiquement *toutes* les régions du cortex cérébral, par l'intermédiaire d'une région plus petite connue sous le nom de cortex entorhinal (figure 10-2). Ces entrées circulent, à travers l'hippocampe, le long d'une séquence de trois synapses successives. Ayant traversé ces structures, les signaux retournent en formant une boucle vers le cortex entorhinal, puis sont relayés par des fibres réentrantes vers les aires corticales dont ils étaient initialement issus. Indirectement, toutes les cellules appartenant à cette boucle de l'hippocampe sont liées de façon simultanée au mésencéphale et aux aires hédonistes, c'est-à-dire aux aires ayant trait aux valeurs.

A quoi servent tous ces circuits ? Tout incite à penser que l'hippocampe est nécessaire à l'établissement de la mémoire à long terme. Le célèbre H. M. – un malade qui avait subi une ablation de l'hippocampe parce qu'il était atteint d'une forme d'épilepsie qui pouvait être mortelle – présentait un syndrome très net à la suite de l'intervention chirurgicale. Il était capable de se rappeler tous les souvenirs qui remontaient à quelque temps avant son opération, mais incapable de se remémorer les événements qui s'étaient produits peu de temps avant qu'on ne l'interroge à leur sujet. S'il venait à être interrompu au moment où il donnait son adresse, par exemple, il perdait la capacité de s'en souvenir. H. M. était tout à fait conscient et parvenait même à apprendre un certain nombre de séquences motrices. Cependant, la perte de l'hippocampe l'ayant rendu à tout jamais infirme, il ne parvenait plus à emmagasiner à long terme de nouveaux souvenirs, des souvenirs actuels.

Les études chez l'animal ont permis de confirmer le rôle de l'hippocampe dans la transformation des réponses à des tâches à court terme en vue de leur stockage dans la mémoire à long terme. Selon toute vraisemblance, le rôle de cet appendice cortical est

d'aider à ordonner les événements qui ont été catégorisés sur le moment par le cortex, puis d'assurer que ces événements catégorisés entraînent des modifications synaptiques supplémentaires dans le cortex, garantissant ainsi une mémoire à long terme.

Par rapport à la TSGN, tout cela signifie que, bien que les couples de classification et les cartographies globales soient soumis à la sélection des groupes neuronaux à travers les modifications synaptiques, cela ne suffit pas, en soi, à assurer la liaison entre mémoire à court terme et mémoire à long terme. À moins que des organes de la succession tel que l'hippocampe interviennent afin d'ordonner les résultats, de graves défauts de la mémoire s'ensuivront.

Si j'ai alourdi mon propos avec ces détails anatomiques et physiologiques, c'est pour souligner le rôle interactif des différentes structures cérébrales et de la dynamique du cerveau dans la réalisation des diverses fonctions psychologiques. Si les structures et la neuro-anatomie du cervelet, ou quelque chose de semblable, n'étaient pas apparues au cours de l'évolution, nous ne pourrions exécuter des mouvements bien coordonnés, enchaînés et rapides. Par ailleurs, sans les ganglions de la base et leur anatomie spécifique, les animaux auraient été incapables d'orchestrer des symphonies entières de mouvements au sein d'un plan. Enfin, sans les fonctions fournies par l'hippocampe, des pans entiers de catégorisations, allant de celles mémorisées dans l'immédiat à celles emmagasinées pour toujours, ne pourraient pas être reliés. En l'absence d'une telle liaison, la mémoire à long terme n'aurait aucune chance d'être cohérente.

À travers tous ces exemples, nous avons vu que la morphologie et les circuits cérébraux qui se sont développés au cours de l'évolution, modulés par la biochimie des synapses, peuvent fournir de nouvelles fonctions et de nouveaux types de mémoire. À partir d'études effectuées chez les invertébrés et chez les vertébrés « inférieurs », on voit clairement que l'action du système nerveux consiste à régler les fonctions corporelles et les comportements. Des types particuliers de mémoire, fondés sur les modifications synaptiques, existent certainement chez ces animaux. Mais au cours de la vie terrestre (et, pour les précurseurs des hominidés, au cours de la vie dans les arbres), l'environnement mit les êtres vivants face à de nouveaux et immenses défis. L'apparition d'un cortex capable d'effectuer des catégorisations perceptives et d'organes de la succession capables de les ordonner a permis d'utiliser des ensembles beaucoup plus riches de fonctions psychologiques pour affronter des

environnements complexes. Ces structures ont modifié le sens du mot mémoire. Remarquez cependant que les principes de sélection et de réentrée suffisent à eux seuls pour acquérir de nouvelles fonctions de mémoire. Car ce qu'il faut, ce sont de nouvelles structures ou morphologies – de nouveaux arrangements des connexions cérébrales, comme ceux que l'on trouve dans les appendices corticaux.

LES CONCEPTS

Les mêmes principes s'appliquent à la catégorisation proprement dite. De fait, l'incarnation d'aptitudes de catégorisation étendues requiert un autre progrès évolutif, hormis la mémoire en tant que recatégorisation. C'est ce que j'appelle la capacité de former des concepts, et le point de vue que j'ai adopté s'écarte, tout comme dans le cas de la mémoire, de la vision conventionnelle des choses. En effet, l'utilisation du mot « concept » est en général associée au langage, et à des contextes où les choses sont toujours soit vraies, soit fausses. Quant à moi, lorsque j'utilise le mot concept, je fais référence à une faculté qui apparaît, au cours de l'évolution, avant l'acquisition de primitives linguistiques. De quoi s'agit-il ?

Les animaux capables de former des concepts sont capables d'identifier les choses ou les actes et, à partir de cette identification, de régler leur comportement d'une manière plus ou moins générale. Cette reconnaissance doit être relationnelle : elle doit pouvoir relier une catégorisation perceptive à une autre, apparemment sans rapport avec la première, même en l'absence des stimulus qui ont déclenché ces catégorisations. Les relations ainsi captées doivent permettre des réponses à des propriétés générales – telles qu'« objets », « haut/bas », « dedans », et ainsi de suite. Contrairement aux éléments du langage, cependant, les concepts ne sont *pas* conventionnels ou arbitraires, leur développement ne requiert *pas* un rattachement à une communauté linguistique, et ils ne dépendent *pas* d'une présentation séquentielle. Les capacités conceptuelles sont apparues, au cours de l'évolution, bien avant le langage. Et, bien qu'elles dépendent de la perception et de la mémoire, elles sont *construites* par le cerveau à partir d'éléments qui contribuent à ces deux fonctions à la fois.

Il est difficile de savoir quels sont les animaux, en dehors des humains, qui ont des capacités conceptuelles. Il est indubitable, en

tout cas, que tout ce que l'on sait sur les chimpanzés porte à croire que ces animaux en ont. Les chimpanzés généralisent et classifient des relations faisant intervenir aussi bien des objets que des actions. En revanche, il est plus difficile de déterminer quelles sont les capacités conceptuelles des autres animaux, car contrairement à ce qui se passe avec les chimpanzés, notre communication avec eux est très limitée. Le mieux consiste probablement à comparer les structures et les fonctions de leurs aires cérébrales à celles des êtres humains afin d'émettre des hypothèses qui guideront les études ultérieures.

Comment les capacités conceptuelles sont-elles apparues ? La TSGN postule que le développement, au cours de l'évolution, d'aires cérébrales spécialisées est nécessaire à l'apparition de capacités conceptuelles. L'argument en faveur de cette proposition est fondé sur l'idée selon laquelle un simple accroissement du nombre de cartes réentrantes capables d'effectuer des catégorisations perceptives ne suffit pas à rendre compte de la formation de concepts. Les catégorisations conceptuelles sont extrêmement hétérogènes et générales. Les concepts font appel à des mélanges de relations concernant le monde réel, les souvenirs et les comportements passés. Contrairement aux aires cérébrales intervenant dans la perception, celles qui sont chargées de la conceptualisation doivent pouvoir opérer sans entrées directes.

Quelles sont les opérations cérébrales qui donnent lieu à ces propriétés ? La TSGN suggère que, en formant des concepts, le cerveau construit des cartes de ses *propres* activités, et non pas seulement des cartes des stimulus externes, comme c'est le cas pour la perception. D'après la théorie, les aires cérébrales intervenant dans la formation des concepts contiennent des structures qui classent par catégories, distinguent et combinent les diverses activités cérébrales survenant *dans différents types de cartographies globales.* Au lieu de classer par catégories les entrées extérieures provenant des modalités sensorielles, ces structures cérébrales classent des parties des cartographies globales passées par modalité, selon la présence ou l'absence de mouvement, et selon la présence ou l'absence de relations entre catégorisations perceptives (figure 10-2).

Les structures capables d'avoir ce type d'activités se trouvent vraisemblablement dans les cortex frontal, temporal et pariétal du cerveau. Elles doivent représenter *une cartographie des divers types de cartes.* En effet, elles doivent être capables d'activer ou de reconstituer des portions des activités *passées* des cartographies globales

de différents types – par exemple, celles qui concernent les différentes modalités sensorielles. Elles doivent également être capables de les combiner ou de les comparer. Cela signifie que, pour qu'il soit possible de former des concepts, il faut qu'il existe des connexions réentrantes particulières, allant de ces aires corticales de niveau supérieur vers d'autres aires corticales, ainsi que vers l'hippocampe et les ganglions de la base.

Les aires cérébrales donnant naissance à des concepts doivent non seulement être capables de stimuler des parties des cartographies globales passées, mais elles doivent être capables de le faire indépendamment des entrées sensorielles du moment. Elles doivent également pouvoir distinguer les différentes classes de cartographies globales (par exemple, de distinguer celles qui correspondent à des objets de celles qui correspondent à des mouvements). Et elles doivent ensuite être capables de relier entre elles les portions réactivées des cartographies globales et de permettre le stockage à long terme de ces activités. C'est nécessaire parce que la formation des concepts requiert l'intervention de la mémoire.

Le cortex frontal est l'un des principaux exemples de centre conceptuel cérébral. On ne sait pas suffisamment de choses sur l'organisation de ses cartes pour pouvoir affirmer que la formation de concepts dans cette aire corticale exige une cartographie *topographique*, comme c'est le cas de la catégorisation perceptive. Il est vraisemblable que des cartes qui cartographient les types d'activités se déroulant dans d'autres cartes corticales soient requises ; mais, dans les cartes d'ordre supérieur, la topographie n'est peut-être pas aussi importante. Étant donné ses connexions avec les ganglions de la base et le système limbique, et aussi avec l'hippocampe, le cortex frontal établit également des relations favorisant la catégorisation des valeurs et des expériences sensorielles elles-mêmes. Ainsi, la mémoire conceptuelle est affectée par les valeurs – une caractéristique importante lorsqu'il s'agit d'améliorer la capacité de survie.

Grâce à cette définition des concepts, en vertu de laquelle le cerveau classe ses propres activités par catégories (et en particulier ses catégorisations perceptives), il devient possible de voir comment les images et les catégories généralisées pourraient s'incarner. On voit également qu'il est possible de classer des événements dans la catégorie « passés » sans qu'il soit nécessaire de les faire défiler dans des activités cérébrales actuelles, comme c'est le cas pour la mémoire à court terme et pour la succession, établie par l'hippocampe, qui mène à la mémoire à long terme. De plus, on voit comment les

aires de formation de concepts, en restimulant récursivement des portions des cartographies globales qui contiennent des modifications synaptiques antérieures, donnent lieu à des *combinaisons* de relations et de catégories. Aucun ordre logique intrinsèque, aucune catégorisation classique ou programmation explicite préalable ne sont nécessaires. Et pourtant, les modes de formation de concepts décrits ici pourraient être responsables, de façon tout à fait naturelle, de l'établissement des catégories complexes que j'aborde dans la postface. Enfin, puisque la formation des concepts se fonde sur la triade centrale formée par la catégorisation perceptive, la mémoire et l'apprentissage, elle est, de par sa nature même, intentionnelle.

Le fait de considérer la mémoire comme une recatégorisation et les concepts comme les produits d'un cerveau qui classe ses propres activités par catégories va nous fournir les éléments nous permettant d'atteindre notre objectif : donner une description biologique de la conscience. En partant des principes de base de la TSGN, nous avons pu relier la triade fondamentale constituée par la catégorisation perceptive, la mémoire et l'apprentissage à l'apparition de capacités conceptuelles. Remarquez cependant que nous n'avons eu besoin d'aucune hypothèse théorique supplémentaire, mais seulement de supposer qu'il se produit, au cours de l'évolution, des changements dans la morphologie cérébrale qui modifient l'organisation des connexions réentrantes. Il se trouve qu'une autre modification de la connectivité réentrante va également nous permettre de comprendre comment nous sommes devenus des êtres conscients.

Chapitre 11

La conscience ou le présent remémoré

Quelque chose de tout à fait précis se produit lorsque, à un certain état du cerveau correspond un certain « état de connaissance ».

William James

La plupart des gens, si on leur demandait de dire ce qui rend l'esprit aussi profondément différent et étrange, s'en tiendraient probablement, comme Descartes, au solo du moi et répondraient : « La conscience. » Nous sommes maintenant parvenus au point de notre voyage où nous pouvons nous demander si nous ne pourrions mieux faire que de postuler l'existence d'une substance pensante inaccessible à la science des choses étendues.

Ce que la conscience a de désarmant, c'est qu'elle ne semble pas relever du comportement. Elle *est*, tout simplement – toujours là, avec ses modes et ses objets multiples et simultanés, toujours inéluctablement nôtre. C'est un processus, et un processus difficile à analyser. Nous savons ce qu'elle est pour nous-mêmes, mais nous ne pouvons juger de son existence chez les autres que par analogie. Comme le disait James, il s'agit de quelque chose dont le sens « nous est connu tant que personne ne nous demande de le définir ».

De fait, la meilleure façon de la définir consiste tout d'abord à considérer certaines de ses propriétés. Bien sûr, on est tenté de se laisser aller à des définitions circulaires et de parler de « connaissance de soi ». Considérons en particulier ce que j'appelle les propriétés « jamesiennes » de la conscience (d'après James, qui les analysa) : elle est personnelle (elle appartient à l'individu, au moi) ; elle est changeante mais continue ; elle a affaire à des objets qui sont indépendants d'elle ; elle est sélective dans le temps – autrement dit, elle n'épuise pas tous les aspects des objets auxquels elle a affaire.

La conscience est intentionnelle ; on est conscient des choses ou des événements eux-mêmes, ou encore de choses ou d'événements les concernant. Elle est aussi, dans une certaine mesure, liée à la volonté. Certains psychologues ont d'ailleurs suggéré que la cons-

cience est marquée par la présence d'images mentales et par leur utilisation dans la régulation des comportements. Mais la conscience n'est *pas* une simple copie de l'expérience (un « miroir de la réalité »), et il existe beaucoup de comportements pour lesquels elle n'est même pas nécessaire. Certaines formes d'apprentissage, certains processus conceptuels, et même certaines formes de raisonnement par analogie s'en passent complètement.

Quant à moi, je distingue – et il s'agit là d'une distinction essentielle – la conscience primaire de la conscience d'ordre supérieur. La conscience primaire est l'état qui permet de se rendre compte de la présence des choses dans le monde – d'avoir des images mentales dans le présent. Mais elle ne s'accompagne pas d'un sens de la personne, avec son passé et son présent. C'est ce type de conscience dont on peut présumer l'existence chez certains animaux dépourvus de capacités linguistiques et sémantiques (je reviendrai plus loin sur les animaux dont il pourrait s'agir). En revanche, la conscience d'ordre supérieur fait appel à la reconnaissance par un sujet pensant de ses propres actes et affects. Elle incarne un modèle personnel, un modèle du passé et du futur aussi bien que du présent. Elle dénote une conscience directe – la conscience immédiate, non réfléchie de l'existence d'épisodes mentaux, sans aucune intervention des organes récepteurs ou sensoriels. C'est cette aptitude que nous autres humains possédons en plus de la conscience primaire. Nous sommes *conscients d'être conscients.*

Le mot « conscience » possède également d'autres connotations. Celles-ci sont mises en évidence, par exemple, par les critères utilisés par les cliniciens pour déterminer si un malade qui a subi un traumatisme est ou non « conscient » – critères qui ont trait à l'éveil, à l'orientation, à la conscience de soi, et au contrôle des motivations. Et, lorsque l'acuité perceptive et la capacité de la mémoire sont diminuées, les médecins parlent d'« obscurcissement » de la conscience. Dans les cas extrêmes, dus à la maladie, les propriétés jamesiennes, les « envolées et les perchoirs de la conscience » deviennent aléatoires, automatiques, répétitifs, et ne donnent plus de signe d'une introspection ou d'un éveil de l'attention face à la nouveauté. Dans le cas le plus extrême de tous – il n'y a plus rien, absolument plus rien à signaler.

Les hypothèses sur la conscience, en particulier celles des philosophes, sont innombrables. Mais la plupart d'entre elles ne forment pas ce que nous pourrions appeler des théories scientifiques bâties sur des principes, fondées sur l'observation et associées aux

fonctions du cerveau et du corps. Plusieurs théories fonctionnalistes et computationnelles de la conscience (voir la postface) ont récemment été formulées. Elles sont en général de l'un des deux types suivants : la conscience est supposée efficace ou bien considérée comme un épiphénomène. Dans le premier cas, la conscience est assimilée aux programmes superviseurs des systèmes informatiques et dans le second, à un sous-produit fascinant, mais plus ou moins inutile, des calculs.

Cependant, aucune de ces notions ne fait directement appel à la biologie ou à la nature de l'incarnation de la conscience. Or il est évident que cela est indispensable à toute théorie de la conscience fondée sur l'évolution. Une théorie de ce genre doit en effet proposer des modèles *neuronaux* explicites permettant d'expliquer comment apparaît la conscience. Elle doit nécessairement expliquer comment la conscience apparaît au cours de l'évolution et du développement. Elle doit relier la conscience aux autres aspects de la vie mentale, tels la formation des concepts, la mémoire et le langage. Et elle doit décrire des tests permettant de confronter, de façon très rigoureuse, les modèles qu'elle propose aux données neurobiologiques. Ces tests devront être effectués de préférence au moyen d'expériences réelles, et au moins à travers ce que l'on appelle des *Gedanken experimenten* – des expériences de pensée. Dans ces dernières, toutes les propriétés postulées devront être parfaitement compatibles avec les observations scientifiques actuellement connues, quel que soit le domaine de recherches dont elles sont issues, mais, surtout, avec celles de la neurobiologie.

Dans l'état actuel des choses, il s'agit là d'un programme très ambitieux, car les analyses biologiques de la conscience ressemblent à celles des premiers événements cosmologiques : on sait d'emblée que certaines manipulations et observations sont tout simplement irréalisables. Dans ces conditions, on doit prendre soin d'expliciter dans leur entier les hypothèses sous-jacentes aux théories proposées. Je vais donc en expliciter trois, qui font toutes partie des fondements de ma théorie de la conscience. Deux d'entre elles sont simples, la troisième est un peu délicate. Je les appelle hypothèse physique, hypothèse évolutionniste et hypothèse des sensations [1] (c'est elle la plus délicate). Je dois énoncer clairement ces hypothèses dès le

1. L'expression utilisée de l'auteur est *qualia hypothesis, les qualia* sont les qualités sensibles des choses, par opposition à leurs caractéristiques mesurables, les *quanta* (N.d.T.).

départ pour éviter de tomber dans certains pièges – notamment dans le cartésianisme, le panpsychisme ou encore dans le bourbier cognitiviste-objectiviste dont il est question dans la postface.

Selon l'hypothèse physique, si les lois de la physique sont respectées, les âmes et les fantômes sont exclus. La description du monde donnée par la physique moderne constitue le fondement adéquat, mais pas tout à fait suffisant, de toute théorie de la conscience. En effet, la théorie moderne des champs quantiques fournit une description d'un ensemble de propriétés formelles de la matière et de l'énergie à toutes les échelles (voir la figure P-1), mais elle ne contient pas de théorie de l'intentionnalité ni de théorie des noms pour les objets macroscopiques, car elle n'en a pas besoin. Ce que j'entends lorsque j'affirme que la physique n'est qu'adéquate, c'est que je ne permettrai à aucun épouvantail – gravitation quantique, action à distance, superphysique (voir la postface) – d'entrer dans cette théorie de la conscience.

L'hypothèse évolutionniste, elle aussi, est assez simple. Elle consiste à dire que la conscience est une propriété phénotypique apparue à un moment donné de l'évolution des espèces. Avant cela, elle n'existait pas. Cette hypothèse suggère que l'acquisition d'une conscience a directement conféré un avantage adaptatif aux individus qui en avaient une ou bien a permis l'apparition d'autres traits qui, eux, amélioraient l'adaptation. Elle suggère également que la conscience est efficace – c'est-à-dire qu'elle *n'est pas* un épiphénomène (« rien que la couleur rouge du métal en fusion », alors que ce qui compte, c'est qu'on puisse le couler).

Avec la troisième hypothèse, cependant, nous abordons des questions plus subtiles. Il s'agit de questions de méthodologie, que pose la façon curieuse dont la conscience se manifeste. Pour expliquer la difficulté, je dois d'abord faire un détour par les propriétés phénoménales ou senties, c'est-à-dire par les sensations.

Les sensations correspondent à l'ensemble des expériences personnelles ou subjectives, des sentiments et des impressions qui accompagnent l'état de conscience. Ce sont des états phénoménaux qui constituent le « comment nous voyons les choses » en tant qu'êtres humains. Par exemple, la « couleur rouge » d'un objet rouge est une sensation. Les sensations sont les parties discernables d'une scène mentale qui néanmoins présente une unité d'ensemble. Elles peuvent varier en intensité et en clarté, elles peuvent aller de la « sensation pure » à des discriminants particulièrement raffinés. Elles peuvent être très précises lorsqu'elles accompagnent des expé-

riences perceptives ; en l'absence de perception, elles peuvent être plus ou moins diffuses, mais néanmoins discernables en tant que « visuelles », « auditives », et ainsi de suite. En général, dans l'état de veille normal, les sensations s'accompagnent d'un sens de continuité spatio-temporelle. Souvent, la scène phénoménale s'accompagne de sentiments ou d'émotions, plus ou moins ténus. Cependant, la *séquence* de sensations que l'on éprouve effectivement est extrêmement personnelle, car elle repose sur une série de moments de l'histoire personnelle ou de l'expérience immédiate de chacun.

Étant donné que les sensations ne sont directement éprouvées que par chaque individu séparément, notre difficulté méthodologique est évidente. *Nous ne pouvons construire une psychologie phénoménale qui puisse être partagée avec d'autres à l'instar de la physique.* Les sensations directement éprouvées par un individu donné ne peuvent pas être totalement partagées par un autre individu jouant le rôle d'observateur. Tout individu peut décrire son expérience à un observateur, mais cette description sera toujours partielle, imprécise et relative au contexte personnel de l'individu. Non seulement les sensations sont fugaces, mais les interventions destinées à les sonder peuvent les modifier de manière imprévisible. De plus, beaucoup de processus conscients et non conscients affectent l'expérience subjective de chaque personne. Par conséquent, chacun est susceptible de former sa propre théorie concernant la totalité de son expérience *individuelle* consciente, mais ces théories ne seront jamais des théories scientifiques, puisque les autres observateurs ne disposent pas de sujets témoins adéquats pour contrôler les expériences.

Le paradoxe est poignant : pour faire de la physique, je fais appel à ma vie consciente, à mes perceptions, à mes sensations. Mais dans la communication intersubjective, je les exclus de ma description, certain que je suis que mes confrères observateurs, munis chacun de leur vie consciente personnelle, pourront effectuer les manipulations prescrites pour arriver à des résultats expérimentaux comparables. Et lorsque, pour une raison ou pour une autre, les sensations influent effectivement sur les interprétations, on modifie le protocole expérimental afin de proscrire de tels effets : on place l'esprit hors de la nature.

Mais si nous voulons étudier la conscience, nous ne pouvons pas négliger les sensations. Le dilemme réside alors dans le fait que l'expérience phénoménale est une question qui se traite à la première personne, ce qui semble, à première vue, interdire la for-

mulation d'une description entièrement objective ou causale. La situation est-elle pour autant totalement désespérée ?

Je ne le pense pas. Mais quelles sont les alternatives qui nous restent si nous voulons parvenir à analyser scientifiquement la conscience ? L'une de ces alternatives, qui ne semble absolument pas viable, consiste à ignorer complètement l'existence de sensations. On pourrait alors formuler une théorie de la conscience qui viserait *par ses seules descriptions* à transmettre à un hypothétique observateur, « libre de toute sensation », ce que signifie sentir la chaleur, voir du vert, et ainsi de suite. En d'autres termes, il s'agit d'essayer de formuler une théorie fondée sur une sorte de vision extérieure de la conscience. Mais aucune théorie *scientifique*, quelle qu'elle soit, n'est possible sans que l'on suppose d'emblée que les observateurs ont aussi bien des sensations que des perceptions. Supposer le contraire serait commettre les mêmes erreurs que les théories qui tentent de plaquer des formulations syntaxiques sur des interprétations objectivistes : elles négligent le fait que l'incarnation est productrice de sens (voir la postface). Aucun observateur scientifique n'est exempt de sensations.

Si nous excluons cette voie, quel autre recours nous reste-t-il ? Je pense qu'il nous en reste un, qui repose sur le fait que les êtres humains occupent une position privilégiée. Bien que nous ne soyons pas nécessairement les seuls animaux conscients, nous sommes, peut-être avec les chimpanzés, les seuls animaux conscients d'eux-mêmes. Nous sommes les seuls animaux capables de parler, de modéliser le monde en dépassant l'instant présent, de décrire nos états phénoménaux, de les étudier et de les corréler avec les résultats de la physique et de la biologie.

Ces aptitudes suggèrent une approche possible du problème des sensations. Pour fonder une théorie de la conscience, il est raisonnable de *supposer* que, tout comme elles existent chez nous-mêmes, elles existent également chez les autres êtres humains conscients – qu'ils soient considérés comme des observateurs ou comme des sujets. (Le fait que les sensations puissent varier d'un observateur à un autre importe peu ; ce qui compte, c'est que chaque observateur en ait.) Nous pouvons alors considérer les êtres humains comme le meilleur référent canonique pour l'étude de la conscience. C'est légitime parce que les descriptions subjectives fournies par les êtres humains (y compris celles qui portent sur des sensations), leurs actions, ainsi que la structure et la fonction de leur cerveau, *peuvent toutes être corrélées*. Une fois que nous aurons bâti une théorie

fondée sur l'hypothèse que les êtres humains ont des sensations, nous pourrons examiner à nouveau certaines des propriétés de ces sensations d'après ces corrélations. C'est notre capacité de décrire et de corréler les choses tout en éprouvant individuellement des sensations en provenance d'elles qui nous donne la possibilité d'étudier scientifiquement la conscience.

L'*hypothèse des sensations* établit une distinction entre la conscience d'ordre supérieur et la conscience primaire. La conscience d'ordre supérieur s'appuie sur l'existence d'une conscience directe, immédiate, chez un être humain possédant un langage et une vie subjective susceptible d'être décrite. La conscience primaire peut être constituée d'expériences phénoménales telles que des images mentales, mais elle est limitée à un intervalle de temps situé autour du présent mesurable, elle est dépourvue de concepts de soi, de passé et de futur, et inaccessible à l'auto-description directe et individuelle. Par conséquent, les êtres qui ne possèdent qu'une conscience primaire sont incapables de construire des théories de la conscience – même fausses !

Un programme de recherche bâti sur les hypothèses que je viens de décrire présente évidemment un certain nombre de difficultés intrinsèques. En effet, nous devons tout d'abord construire un modèle de la conscience primaire, puis nous appuyer sur lui pour en construire un autre, représentant la conscience d'ordre supérieur, et ensuite vérifier tous les liens existant entre chacun de ces modèles et l'expérience phénoménale humaine. Pour être compatible avec l'hypothèse évolutionniste, cette façon de procéder devra permettre d'expliquer comment sont apparues la conscience primaire, puis la conscience d'ordre supérieur, au cours de l'évolution. L'ordre des différentes étapes de ce programme expérimental (qui, d'après l'hypothèse des sensations, doit être fondé sur des corrélations obtenues principalement chez des sujets humains) doit donc être exactement l'inverse de celui du programme théorique, qui doit commencer par s'intéresser aux précurseurs des humains au cours de l'évolution.

J'espère que la raison pour laquelle une théorie biologique fondée sur nos trois hypothèses ne peut pas adopter un point de vue extérieur est à présent devenue claire. En tant que scientifiques, nous ne pouvons attendre d'une théorie de la conscience que, par une description linguistique, elle fasse comprendre intuitivement à un hypothétique animal incapable de toute sensation ce qu'est une sensation. Pour entretenir la communication intersubjective et éta-

blir des corrélations scientifiques – il s'agit là d'une activité humaine –, nous devons supposer l'existence de sensations, car celles-ci ne peuvent être déduites d'aucune théorie en tant qu'expériences. Cela ne signifie pas, cependant, qu'il est impossible de distinguer théoriquement les sensations d'après leur modalité, leur intensité, leur continuité ou leurs propriétés temporelles et spatiales. Cela ne signifie pas non plus que, l'hypothèse les concernant étant posée, nous ne puissions pas nous intéresser aux mécanismes qui donnent naissance aux sensations. Notre comparaison avec la cosmologie vient, encore une fois, à propos ; en effet, nous pouvons demander à la physique moderne d'expliquer certains aspects de la cosmologie des premiers temps de l'univers, qui sont compatibles avec ce que nous dit la théorie physique moderne. Mais nous ne pouvons pas demander à une théorie physique de donner une réponse satisfaisante à la question, posée par Gottfried Leibniz, de savoir pourquoi il existe quelque choses plutôt que rien.

Comme nous le verrons après avoir considéré des modèles de la conscience primaire et de la conscience d'ordre supérieur, il peut être utile de considérer les sensations comme des formes de catégorisation d'ordre supérieur, comme des relations susceptibles d'être d'abord communiquées au moi, et ensuite – quoique de façon un peu moins satisfaisante – aux individus ayant un équipement mental semblable. Je sais qu'une déclaration aussi austère ne peut être qu'insatisfaisante. Mais au lieu de m'attarder ici sur ce sujet, je préfère commencer par décrire un modèle de la conscience primaire fondé sur nos trois hypothèses et compatible avec ce que l'on sait sur la structure et la fonction du cerveau. Parmi les éléments de ce modèle se trouvent notamment plusieurs des systèmes déjà décrits, à savoir ceux qui donnent naissance aux valeurs, à la catégorisation perceptive et conceptuelle, et à la mémoire. En outre, la dynamique du modèle dépend d'un type particulier de circuit réentrant. C'est pour cela que j'ai tenu à décrire abondamment toutes ces questions au cours des chapitres précédents. (Pour l'instant, je vais mettre les sensations de côté, mais j'y reviendrai ultérieurement, lorsque j'aborderai la conscience d'ordre supérieur.)

LA CONSCIENCE PRIMAIRE

Le modèle que j'ai proposé comporte un certain nombre de parties. (Feriez-vous confiance à un modèle de la conscience qui n'en

aurait qu'une ?) Mais avant de décrire leurs interactions, et pour qu'ensuite l'explication de ces interactions soit plus claire, je voudrais présenter chacune d'elles. Il existe, en gros, deux types de systèmes, dans le système nerveux, qui sont importants pour comprendre comment la conscience est apparue. Bien qu'ils soient tous deux composés de neurones, ces systèmes possèdent une organisation très différente. Le premier est l'ensemble formé par le tronc cérébral et le système limbique (hédoniste) : il a trait à l'appétit, aux comportements sexuels et consommatoires, et aux stratégies de défense mises en place au cours de l'évolution. Il s'agit d'un système de valeurs relié à un grand nombre d'organes du corps, au système endocrine et au système neuro-végétatif. Cet ensemble règle les rythmes cardiaque et respiratoire, la transpiration, les fonctions digestives et autres, ainsi que les cycles corporels associés au sommeil et à l'activité sexuelle. Vous ne serez certainement pas surpris d'apprendre que les circuits de l'ensemble tronc cérébral-système limbique forment souvent des boucles, qu'ils réagissent relativement lentement (le temps de réaction pouvant aller de quelques secondes à quelques mois) et qu'ils ne comportent pas de cartes détaillées. Ils ont été sélectionnés, au cours de l'évolution, de façon à correspondre au corps, et non à s'ajuster aux multiples signaux imprévus provenant du monde extérieur. Ces systèmes sont apparus tôt au cours de l'évolution afin de prendre soin des fonctions corporelles ; ce sont des systèmes internes.

Le second type majeur d'organisation dans le système nerveux est tout à fait différent. Il s'agit du système dit thalamo-cortical. (Le thalamus, l'une des structures centrales du cerveau, est composé de nombreux noyaux chargés d'acheminer les signaux sensoriels et d'autres signaux cérébraux vers le cortex.) Le système thalamo-cortical résulte de l'action conjointe du thalamus et du cortex. Il est apparu au cours de l'évolution pour permettre de recevoir les signaux provenant des couches de récepteurs sensoriels et d'envoyer des signaux aux muscles volontaires. Ses réactions sont très rapides (entre quelques millisecondes et quelques secondes), bien que ses connexions synaptiques subissent certaines modifications qui durent toute la vie. Sa principale structure, le cortex cérébral, est organisée selon un ensemble de cartes, qui reçoivent des entrées provenant du monde extérieur via le thalamus. Contrairement au système tronc cérébral-système limbique, il comporte, plutôt que des boucles, des structures locales stratifiées fortement interconnectées, dont les connexions sont massivement réentrantes. À beaucoup d'endroits,

celles-ci sont organisées topographiquement (voir la figure 9-2). Le cortex cérébral est une structure adaptée à la réception de séries de signaux denses et rapides arrivant simultanément du monde extérieur par l'intermédiaire de nombreuses modalités sensorielles – la vue, le toucher, le goût, l'odorat, l'ouïe, la proprioception (le sens de la position des articulations). En termes d'évolution, il est apparu plus tard que le système tronc cérébral-système limbique, afin de permettre des comportements moteurs de plus en plus élaborés et la catégorisation des événements extérieurs. Les appendices corticaux – le cervelet, les ganglions de la base et l'hippocampe (voir la figure 10-2) – sont apparus en même temps que le cortex pour traiter des aspects temporels et spatiaux, gérant la succession à la fois dans les mouvements réels et dans la mémoire.

Les deux systèmes – tronc cérébral-système limbique et thalamocortical – furent reliés l'un à l'autre durant l'évolution. En effet, le système cortical, d'évolution plus tardive, permettait un apprentissage adaptatif dans des environnements de plus en plus complexes. Et comme ce comportement d'apprentissage était clairement sélectionné de façon à satisfaire les besoins physiologiques et les valeurs issues de l'ensemble tronc cérébral-système limbique, d'évolution plus précoce, il fallait que cet ensemble et le système cortical soient reliés de telle sorte que leurs activités puissent s'ajuster l'une à l'autre. De fait, un tel ajustement est un aspect essentiel de l'apprentissage. Et, puisque le cortex s'occupe de la catégorisation du monde et l'ensemble tronc cérébral-système limbique des valeurs (ou de leur réglage vis-à-vis des structures physiologiques sélectionnées par l'évolution), il s'ensuit que l'apprentissage pourrait expliquer comment la catégorisation, qui a lieu en présence de certaines valeurs, entraîne des modifications adaptatives du comportement qui satisfont ces valeurs.

Il est certain que même des animaux qui ne font preuve d'aucun comportement conscient sont capables d'apprendre. Cependant, chez certaines espèces animales possédant un cortex cérébral, les catégorisations portant sur différentes parties du monde qui n'entretiennent pas de relations causales entre elles peuvent être corrélées et jointes pour obtenir une *scène*. Par le mot scène, j'entends un ensemble, ordonné dans l'espace et dans le temps, de catégorisations d'événements, familiers ou non, *non nécessairement reliés entre eux de façon physique ou causale*. L'avantage fourni par la faculté de construire une scène est dû au fait que les événements qui, dans le passé, ont pu être importants pour l'apprentissage de l'animal

pourront ainsi être reliés à de nouveaux événements, et ce même s'ils n'y sont absolument pas reliés de façon causale dans le monde extérieur. Et surtout, cette relation pourra s'établir de façon à satisfaire les systèmes de valeurs individuels de l'animal. De cette manière, le fait qu'un événement soit ou non saillant est déterminé non seulement par sa position et son énergie dans le monde physique, mais aussi par la valeur relative qui lui a été accordée au cours de l'histoire individuelle passée de l'animal, à l'issue de l'apprentissage.

C'est le développement, au cours de l'évolution, de la capacité de créer des scènes qui a entraîné l'apparition de la conscience primaire. Mais bien évidemment, si cette conscience a pu survivre, c'est certainement parce qu'elle a permis une meilleure adaptation. Mais avant de voir en quoi, examinons le modèle lui-même.

Selon ce modèle, l'apparition de la conscience primaire dépend de l'évolution de trois fonctions. Deux de ces produits de l'évolution sont nécessaires à l'apparition de la conscience, mais ils ne sont pas suffisants. La première fonction est un système cortical permettant aux fonctions conceptuelles, lorsqu'elles sont apparues, d'être fortement liées au système limbique, entraînant ainsi l'élargissement des capacités d'apprentissage déjà existantes. La seconde fonction est une nouvelle sorte de mémoire, fondée sur cette liaison. Contrairement au système de catégorisation perceptive, ce système de mémoire conceptuelle est capable de *classer par catégories les réponses des différents systèmes cérébraux* qui effectuent des catégorisations perceptives, et de le faire conformément aux exigences des systèmes de valeurs de l'ensemble tronc cérébral-système limbique. Cette mémoire des « valeurs-catégories » permet de produire des réponses conceptuelles en fonction des interactions *réciproques* du système thalamo-cortical et de l'ensemble tronc cérébral-système limbique.

C'est un troisième produit essentiel de l'évolution qui fournit les moyens suffisants à l'apparition de la conscience primaire. Il s'agit d'un circuit réentrant particulier qui est apparu, au cours de l'évolution, comme un nouveau composant neuro-anatomique. Il permet à la mémoire des valeurs-catégories et aux cartographies globales en cours, qui traitent de la catégorisation perceptive en temps réel, d'échanger continuellement des signaux de façon réentrante. Il est vrai que, même en l'absence de ces nouvelles connexions réentrantes, un animal pourra effectuer des catégorisations perceptives dans diverses modalités sensorielles, et même acquérir une mémoire

conceptuelle des valeurs-catégories. Cependant, cet animal ne pourra pas relier les événements perceptifs pour obtenir une scène en cours. En revanche, avec l'apparition de nouveaux circuits réentrants dans chaque modalité, *une catégorisation conceptuelle des perceptions simultanées* peut avoir lieu *avant* que ces signaux perceptifs n'aient laissé une empreinte durable dans cette mémoire des valeurs-catégories. C'est cette interaction entre un type particulier de mémoire et la catégorisation perceptive qui donne naissance à la conscience primaire. Une fois les circuits réentrants adéquats en place dans le cerveau, ce « processus de *bootstrapping* » [2] a lieu simultanément et en parallèle dans toutes les modalités sensorielles, permettant ainsi la construction d'une scène complexe. La cohérence de cette scène est coordonnée par la mémoire conceptuelle des valeurs-catégories, même si les différentes catégorisations perceptives qui y contribuent sont indépendantes du point de vue causal.

J'utilise le mot « scène » de manière à transmettre l'idée selon laquelle les réponses à des événements plus ou moins contemporains survenant dans le monde sont reliées par un ensemble de processus réentrants. Nous autres humains, qui possédons une conscience d'ordre supérieur, considérons la conscience primaire comme un « tableau », ou une « image mentale », d'événements en cours classés par catégories. Mais, comme nous le verrons lorsque nous analyserons la conscience d'ordre supérieur, le cerveau ne comporte en fait aucune image, aucun dessin. L'« image » que nous croyons voir n'est qu'une *corrélation* entre différents types de catégorisations.

Résumons : un processus d'« auto-catégorisation » conceptuelle se déroule dans le cerveau. Le cerveau construit des auto-catégories en ajustant les catégories perceptives passées aux signaux provenant des systèmes de valeurs, processus qui est à la charge de systèmes corticaux capables d'assurer des fonctions conceptuelles. Ensuite, ce système de valeurs-catégories interagit, via des connexions réentrantes, avec les aires cérébrales chargées de la catégorisation perceptive en cours des événements et des signaux provenant du monde extérieur. L'expérience perceptive (phénoménale) naît des corrélations établies par une mémoire conceptuelle sur un ensemble de catégorisations perceptives en cours. Autrement dit, la conscience primaire est une sorte de « présent remémoré ».

Ces notions sont illustrées sur la figure 11-1. Et, bien que le

2. Processus par lequel un système « construit » des structures utiles plus puissantes que celles qui étaient initialement présentes (N.d.T.).

diagramme ne reflète nullement la complexité réelle des circuits neuronaux en jeu, il permet néanmoins de mettre en valeur un certain nombre d'éléments. Le premier concerne ce que l'on peut appeler les composants du soi et du non-soi (par le soi, j'entends l'individu biologique unique, et non le moi « humain » issu des interactions sociales). Le soi, c'est-à-dire les systèmes internes, naît de l'interaction des systèmes limbique et cortical. C'est cela qui le distingue des systèmes relatifs au monde extérieur, qui sont strictement corticaux.

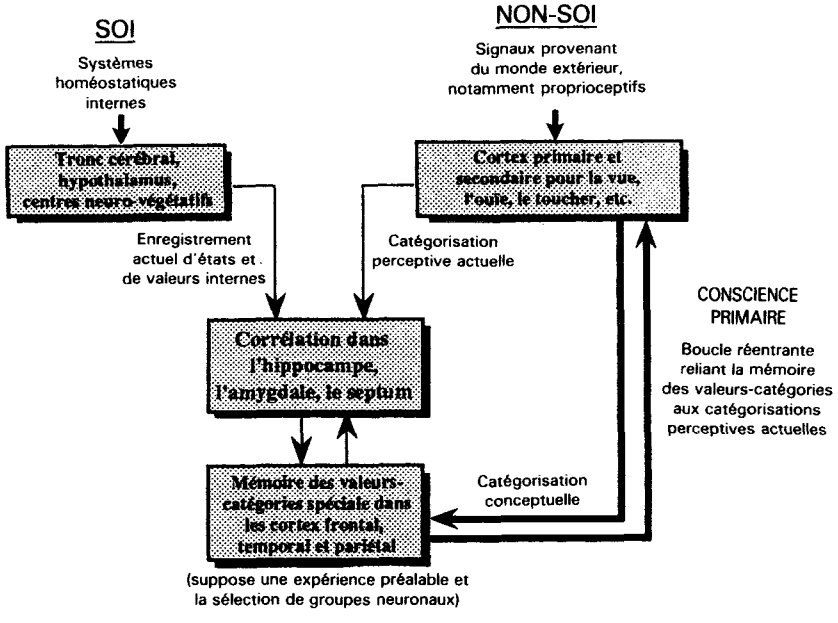


FIGURE 11-1

Un modèle de la conscience primaire. Les signaux passés reliés à des valeurs (définies par des systèmes de contrôle internes) et les signaux provenant du monde extérieur et classés par catégories sont corrélés et conduisent à l'apparition d'une mémoire dans les aires conceptuelles. Cette mémoire, qui est capable d'effectuer des catégorisations conceptuelles, est reliée par des voies réentrantes aux catégorisations perceptives actuelles *des signaux venant de l'extérieur (traits gras). Cela donne lieu à la conscience primaire. Lorsque la même chose se produit pour de nombreuses modalités (vue, toucher, etc.), la conscience primaire est celle d'une « scène » composée d'objets et d'événements, dont certains ne sont pas reliés de façon causale. Malgré tout, un animal doué de conscience primaire est capable de relier ces objets et événements grâce à sa mémoire, via ses expériences antérieures, chargées de valeurs.*

Le second élément concerne la formation de la mémoire des valeurs-catégories. Cette mémoire conceptuelle dépend des interactions constantes entre les systèmes du soi et les systèmes du monde extérieur. Le troisième élément a trait à la réalisation, en temps

réel et *en parallèle*, des catégorisations perceptives pour *chacune* des modalités sensorielles, via le système cortical et notamment les organes de la succession. Et le dernier élément, qui est essentiel, préfigure l'apparition de la conscience primaire : des scènes corrélées résultent de l'activité des connexions réentrantes existant entre les systèmes corticaux qui interviennent dans la mémoire conceptuelle des valeurs-catégories et les systèmes thalamo-corticaux qui interviennent dans les catégorisations perceptives en cours dans toutes les modalités sensorielles.

Vous remarquerez que la conscience primaire telle que je l'ai caractérisée possède les propriétés jamesiennes essentielles : elle est individuelle (les systèmes du soi y contribuent), elle est continue mais changeante (elle change à mesure que les signaux de l'intérieur et ceux de l'extérieur évoluent) et elle est intentionnelle (elle fait nécessairement référence à des signaux internes ou externes dérivés tour à tour des objets et des événements). Si l'on reproduisait la figure 11-1 afin de représenter une suite d'instants, elle servirait à mettre en valeur ces propriétés jamesiennes de la conscience primaire et le type de *bootstrapping* perceptif auquel elle correspond. Les propriétés jamesiennes mettent en valeur le flux de la conscience, son « avant » et son « après ». Au cours du processus conscient, la catégorisation perceptive en cours, qui n'est influencée par aucune valeur, interagit avec une mémoire dominée par les valeurs. Et cela se produit *avant* que les événements perceptifs ne viennent contribuer davantage à la modification de cette mémoire. Ainsi, quand ils finissent par effectivement y contribuer, ils ont en général cessé d'appartenir au présent spécieux ou remémoré ; autrement dit, ils ne se trouvent plus dans la conscience primaire.

Quelle est la valeur d'un tel système du point de vue de l'évolution ? À l'évidence, si cette description biologique est exacte, la conscience primaire est nécessairement efficace ; autrement dit, il ne s'agit pas simplement d'un épiphénomène. En fait, selon la TSCN, la conscience primaire contribue à abstraire et à structurer les modifications complexes qui surviennent dans un environnement faisant intervenir des signaux multiples et parallèles. Et même si certains de ces signaux n'ont aucun lien de causalité direct entre eux dans le monde extérieur, ils peuvent constituer, *pour l'animal*, des indicateurs significatifs d'un danger ou d'une récompense. Cela est dû au fait que la conscience primaire relie leurs caractéristiques en fonction de ce qui compte pour l'animal, ce qui à son tour dépend de l'histoire passée et des valeurs de cet animal.

La conscience primaire fournit un moyen de mettre en relation les signaux d'entrée reçus par un individu sur le moment avec ses actes et avec des récompenses passés. En présentant une scène corrélée, elle fournit un moyen de guider l'attention au cours de l'accomplissement séquentiel de tâches d'apprentissage complexes. Elle fournit également un moyen efficace de corriger les erreurs. Il est vrai que toutes ces tâches pourraient être accomplies sans qu'il soit nécessaire de construire une scène. Mais il semble raisonnable de penser qu'un animal possédant une conscience primaire sera capable d'étendre bien plus rapidement ses capacités d'apprentissage à un bien plus grand nombre de signaux qu'un animal qui en serait dépourvu. La conscience, je le répète, est efficace et susceptible d'améliorer l'adaptation évolutive.

Quoi qu'il en soit, la conscience primaire est nécessaire à l'évolution de la conscience d'ordre supérieur. Mais elle est limitée à un petit intervalle de mémoire situé autour de cette parcelle de temps que j'appelle le présent. Il lui manque une *notion* explicite, un concept de moi personnel, et elle ne permet pas d'intégrer des modèles du passé et du futur dans une scène corrélée. On peut dire que les animaux doués de conscience primaire voient une salle comme si elle était éclairée par un faisceau de lumière : seule la partie sous le faisceau se retrouve explicitement dans le présent remémoré ; le reste est plongé dans l'obscurité. Bien sûr, cela ne veut pas dire que les animaux possédant une conscience primaire ne peuvent pas posséder de mémoire à long terme, ou qu'ils sont incapables d'agir en fonction de cette mémoire. Il est clair qu'ils le peuvent. Mais en général, ils sont incapables d'être conscients de cette mémoire, ou de planifier leur futur à long terme en se fondant sur elle.

Où se trouvent précisément les locus cérébraux qui interviennent dans la conscience primaire ? J'ai déjà écrit ailleurs qu'il était possible que certains circuits situés dans le thalamus, entre le cortex et le thalamus, ou entre les différentes régions corticales, constituent les sites des circuits réentrants clés. Je n'alourdirai donc pas davantage ce chapitre avec des considérations précises sur cette neuroanatomie (voir sur la figure 11-1 les noms des différentes aires concernées). Toutefois, il peut être utile de signaler ici que, comme le montrent les tests cognitifs, certaines lésions cérébrales peuvent conduire à une perte sélective de la reconnaissance explicite *consciente* d'un signal au sein d'un domaine perceptif donné, bien que les

personnes affectées restent quand même capables de le reconnaître implicitement – comme le montrent les tests psychologiques.

Les malades atteints de prosopagnosie – incapacité de reconnaître des visages en tant que tels – à la suite d'un accident vasculaire cérébral illustrent bien ce phénomène. Bien qu'ils n'aient aucune conscience des visages, certains de ces patients, tout en refusant d'admettre qu'ils reconnaissent le visage de leur conjoint, obtiennent des résultats aux tests qui indiquent bien qu'ils possèdent une connaissance fortement discriminante du visage en question. L'agnosie visuelle constitue encore un autre exemple. Les individus atteints de lésions dans le cortex visuel primaire déclarent qu'ils sont aveugles, qu'ils n'ont conscience d'aucune expérience visuelle, mais ils sont néanmoins capables de localiser des objets dans l'espace au cours des tests. Nous analyserons ces questions de manière plus approfondie au chapitre 18, mais je tenais à les évoquer ici pour mettre en évidence le fait qu'il est possible d'en rendre compte en supposant l'existence de perturbations (au sein des domaines perceptifs correspondants) dans les boucles réentrantes dont j'ai postulé l'importance pour la conscience primaire (figure 11-1). Mais remettons à plus tard l'analyse des tests de l'état de conscience.

Avant de passer à l'examen de la conscience d'ordre supérieur, quelques mots sur un certain nombre de questions épineuses s'imposent. Voici la première : quels sont les animaux qui possèdent une conscience primaire ? Je ne peux pas vraiment répondre, si ce n'est en prenant comme base le référent humain sur lequel nous nous sommes mis d'accord. En remontant en arrière, nous pouvons être raisonnablement certains (pour des raisons qui deviendront claires par la suite) que les chimpanzés ont une conscience primaire. Et, selon toute vraisemblance, la plupart des mammifères et certains oiseaux aussi, bien que nous ne puissions en déceler la présence que de manière indirecte. Malheureusement, nous ne disposons pour cela que de tests anatomiques ou de tests de comportement (qui ne font appel ni à la communication par signes ni à des descriptions). Quoi qu'il en soit, si les systèmes cérébraux requis par le présent modèle représentent la *seule* voie évolutive vers la conscience primaire, nous pouvons être à peu près certains que les animaux dépourvus de cortex ou d'un quelconque équivalent n'ont pas de conscience primaire. Cependant, il est amusant d'imaginer que si les animaux à sang froid possédant des cortex primitifs en avaient une, elle serait probablement assez restreinte parce que leurs systèmes de valeurs et leur mémoire des valeurs-catégories ne se trou-

veraient pas dans un environnement biochimique suffisamment stable pour assurer adéquatement la liaison avec un système susceptible d'entretenir cette conscience. Donc, les serpents sont peut-être candidats (c'est peu probable, et c'est fonction de la température), mais les homards sont certainement hors course. Si les études à venir confirment cette hypothèse, cela voudra dire que la conscience a environ trois cents millions d'années.

Chapitre 12

Langage et conscience d'ordre supérieur

> *La conscience humaine tend toujours à un langage et même à un style. Prendre conscience, c'est prendre forme. Même dans les étages inférieurs à la zone de la définition et de la clarté, il existe encore des formes, des mesures, des rapports. Le propre de l'esprit, c'est de se décrire constamment lui-même.*
>
> Henri Focillon

Au cours des deux derniers chapitres nous avons péniblement parcouru un terrain difficile et bigarré. Si vous êtes assez patient pour supporter encore un bout de chemin avec moi, je pense qu'ensuite, lorsque vous regarderez en arrière, vous y verrez beaucoup plus clair – un « déclic » se produira. Au point où nous en sommes, ce n'est pas encore possible – pour « voir » clairement comment fonctionne la conscience primaire, il est nécessaire d'examiner comment apparaît la conscience d'ordre supérieur et en quoi elle diffère de la conscience primaire.

Curieusement, nous autres humains, avec notre conscience d'ordre supérieur, nous ne parvenons pas à « voir le monde » à travers notre seule conscience primaire. Les créatures qui ne possèdent qu'une conscience primaire se forment des images mentales, mais n'ont pas la capacité de les considérer sous l'angle d'un moi socialement constitué. À l'inverse, les êtres qui possèdent un tel moi du fait de leur conscience d'ordre supérieur en ont *besoin* pour relier chaque image mentale à la suivante afin d'apprécier les mécanismes de la conscience primaire ! Sans une conscience d'ordre supérieur, on perd la puissance descriptive qu'elle autorise. (Je me demande d'ailleurs souvent si ce n'est pas cet abandon que certains mystiques recherchent.)

Ce que nous pouvons faire avant d'aborder la question des origines de la conscience d'ordre supérieur, car je pense que ce sera utile, c'est voir quelles sont, dans les modèles que nous avons proposés, les « fonctions » qui assurent les divers types de catégorisation. La catégorisation perceptive, par exemple, n'est pas consciente et peut être effectuée par des couples de classification,

ou même par des automates. Elle traite *des signaux provenant du monde extérieur* – c'est-à-dire, des signaux provenant des couches et des organes sensoriels. En revanche, la catégorisation conceptuelle fonctionne de l'intérieur du cerveau, elle a besoin de la catégorisation perceptive et de la mémoire, et son substrat, ce sont *les activités de diverses portions des cartographies globales.* Le fait de relier les deux types de catégorisations à l'aide d'une voie réentrante supplémentaire pour chaque modalité sensorielle (c'est-à-dire, une voie en plus de celle qui permet à l'apprentissage conceptuel de se dérouler) donne lieu, dans la conscience primaire, à une scène corrélée ou « image ». Cette image peut être régénérée en partie par la mémoire chez les animaux qui possèdent une conscience primaire, mais elle ne peut pas y être régénérée par référence à une mémoire *symbolique.* J'entends par là une mémoire des symboles et des significations qui leur sont associés. Par conséquent, les animaux n'ayant qu'une conscience primaire sont fortement sous l'emprise de la succession des événements en temps réel.

Est-il possible d'abolir la tyrannie exercée par ce présent remémoré ? La réponse est en gros que c'est possible grâce à l'apparition de nouvelles formes de mémoire symbolique et de nouveaux systèmes au service des communications et des transmissions sociales. Dans sa forme la plus évoluée, cela revient à l'acquisition, au cours de l'évolution, de facultés linguistiques. Et, dans la mesure où l'espèce humaine est la seule qui possède un langage, cela veut donc également dire que la conscience d'ordre supérieur a fleuri dans notre espèce. Mais tout porte à croire que ses origines remontent, du moins en partie, aux chimpanzés. Les deux espèces sont capables de penser – et pas seulement de former des concepts – et les chimpanzés semblent également posséder certains éléments d'un concept de moi. Il est certain qu'à la *base*, la reconnaissance de la relation sujet-prédicat chez les humains implique un embryon de conscience de la différence entre le moi (dans le sens social d'« individualité ») et des entités qualifiées d'extérieures au moi. Et il est certain que les chimpanzés ont des comportements indiquant qu'ils sont capables de faire cette distinction. Mais, comme ils n'ont pas de vrai langage, j'affirme que ce que j'appelle conscience d'ordre supérieur ne peut pas s'épanouir chez eux comme elle s'épanouit chez nous.

La conscience d'ordre supérieur exige, à l'évidence, que les structures qui servent à la conscience primaire continuent de fonctionner. De plus, elle fait intervenir la capacité de bâtir une individualité

fondée sur les interactions sociales, de modéliser le monde en fonction du passé et du futur et d'être directement conscient des choses. Or, sans une mémoire symbolique, ces facultés ne peuvent pas se développer.

Pour retracer la façon dont ces capacités ont pu se développer grâce à l'apparition, au cours de l'évolution, d'une mémoire symbolique, nous devons nous demander comment le langage est apparu et comment il s'acquiert. Par conséquent, nous allons d'abord voir que l'émergence d'un vrai langage requiert l'évolution d'un tractus vocal et de centres cérébraux de production et de compréhension du langage parlé. Ensuite, j'aborderai la question centrale de cet ouvrage : déterminer si les concepts se forment ou non avant l'acquisition du langage. Cela m'amènera à conclure qu'il a probablement fallu qu'un modèle de l'interaction moi/non-moi émerge avant que n'apparaisse un vrai langage parlé.

THÉORIE ÉPIGÉNÉTIQUE DU LANGAGE

Les réflexions exposées jusqu'ici suggèrent qu'un modèle de l'acquisition du langage doit nécessairement reposer sur la conscience primaire. En outre, le développement d'une syntaxe et d'une grammaire suffisamment riches serait peu probable sans l'évolution préalable d'un moyen neuronal de formation de concepts. Si cela se confirme, la raison pour laquelle les ordinateurs sont incapables de traiter des problèmes sémantiques devient évidente. Leur mise en œuvre n'est pas la bonne ; elle ne mène pas à la conscience.

Je postule que, avant l'apparition du langage, le cerveau possédait déjà les bases nécessaires aux capacités sémantiques, puisqu'il était capable de produire des concepts et d'agir en fonction d'eux. L'apparition de riches mémoires conceptuelles chez les primates et, chez les hominidés, de facultés phonatoires ainsi que de régions cérébrales particulières pour la production, l'organisation et la mémorisation des sons du langage, permit alors l'émergence de la conscience d'ordre supérieur. (Je ne compte pas donner ici une description détaillée des systèmes grammaticaux, mais on trouvera des passages sur certains aspects pertinents de la grammaire dans la postface.)

Le langage est un attribut particulier et unique de l'*Homo sapiens*. Pouvons-nous rendre compte de son apparition sans creuser un abîme entre la théorie linguistique et la biologie ? Oui, pourvu que nous le fassions en termes épigénétiques autant que génétiques.

Mais cela suppose que l'on mette de côté tout dispositif d'acquisition du langage génétiquement programmé. Cependant, cela ne signifie pas qu'aucune structure spécialisée n'ait été nécessaire à l'apparition du langage. D'ailleurs, on n'a aucun mal à trouver des preuves de l'existence de structures spécialisées héréditaires associées au langage. Lorsque les hominidés sont devenus bipèdes, des modifications sont intervenues dans la structure de la base de leur crâne (figure 12-1). Une base morphologique a ainsi été fournie à l'évolution d'un élément anatomique propre aux humains : le tractus, ou chambre supralaryngé(e). Ce tractus arrive à maturité chez l'enfant humain lorsque le larynx descend. (Pour que les êtres humains ne s'étouffent pas en mangeant, il faut que leur épiglotte se ferme lorsqu'ils déglutissent. De fait, contrairement aux autres animaux, nous ne pouvons pas articuler des sons tout en avalant sans risquer la mort.) Avec le tractus se sont également développées les cordes vocales et la langue ; le palais et les dents ont été sélectionnés de manière à permettre une meilleure régulation du flux d'air sur les cordes vocales, ce qui à son tour permit la production de sons co-articulés, ou phonèmes.

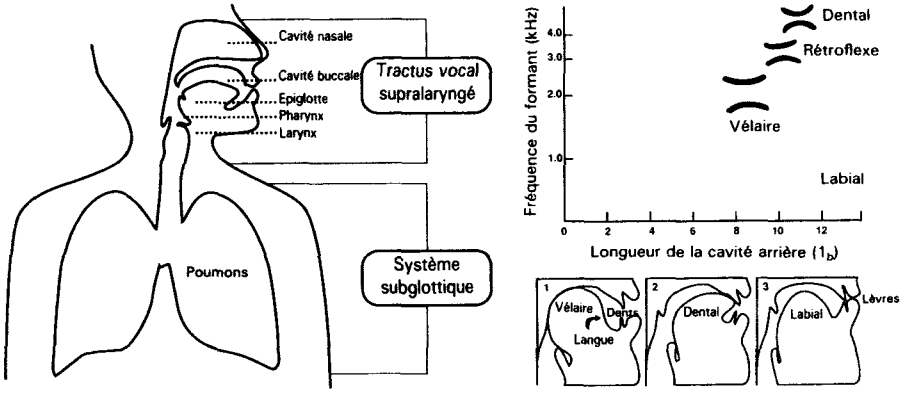


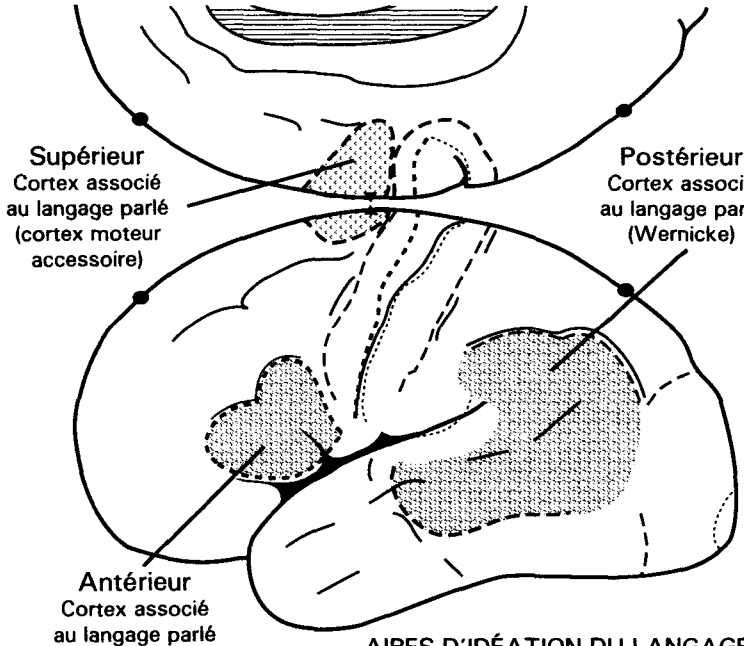
FIGURE 12-1

Le tractus supralaryngé chez l'être humain, l'une des principales bases anatomiques de la parole et de l'évolution du langage. Cette structure complexe fonctionne parce que le larynx (boîte vocale) étant descendu, l'air expiré fait efficacement vibrer les cordes vocales, dont la tension et l'apposition sont à leur tour modifiées grâce à des modifications musculaires extrêmement précises (à gauche). *La modulation exercée par les autres éléments de l'appareil vocal – la langue, les dents, les lèvres, etc. – permet l'émission d'un ensemble de sons articulés* (en haut à droite). *On risque de mourir étouffé si l'épiglotte ne ferme pas les voies respiratoires lorsqu'on déglutit* (en bas à droite). *Ces modifications ne peuvent se produire que si d'autres modifications ont déjà eu lieu à la base du crâne durant l'évolution.*

En même temps ou peu après, des régions particulières du cortex cérébral sont apparues sur le côté gauche : elles sont aujourd'hui connues sous le nom d'aires de Broca et de Wernicke (figure 12-2). Ces régions corticales reliaient les aires auditives, motrices et conceptuelles du cerveau par des connexions réentrantes. À travers ces connexions, les aires de Broca et de Wernicke servaient à coordonner la production et la catégorisation du langage. Et surtout, elles fournissaient un système permettant le développement d'un nouveau type de mémoire, capable d'effectuer des recatégorisations sur les phonèmes (les unités de base du langage) aussi bien que sur l'ordre des phonèmes.

Nous pouvons raisonnablement supposer que la phonétique est apparue dans une communauté linguistique utilisant des phrases primitives (semblables peut-être à celles des sabirs ou pidgins actuels) comme principales unités d'échange. Dans ce type de communautés primitives, les énoncés établissaient des corrélations entre des noms et des objets, marquant ainsi les débuts de la dimension sémantique (figure 12-3). Puis vinrent les verbes. Remarquez que la capacité préexistante de formation de concepts constitua une base indispensable au développement sémantique. Chez les premiers humains, l'organisation pré-syntactique des gestes a peut-être autorisé l'apparition de séquences simples de noms et de verbes. Puis, à mesure que les aires de Broca et de Wernicke se développaient davantage, l'organisation sensori-motrice plus élaborée qui est à la base de la vraie syntaxe est devenue possible.

Selon la théorie d'acquisition du langage que je privilégie, la syntaxe est apparue de façon épigénétique à l'issue d'une séquence précise (figure 12-3). Tout d'abord, les capacités phonétiques ont été reliées, grâce à l'apprentissage, à des concepts et à des gestes, ce qui a permis le développement de capacités sémantiques. Ainsi un lexique est devenu possible, c'est-à-dire des mots et des expressions ayant un sens. La syntaxe est apparue, permettant de faire le lien entre l'apprentissage conceptuel préexistant et l'apprentissage lexical. Une idée semblable a été proposée par Steven Pinker et autres dans le cadre de la grammaire mise au point par Joan Bresnan, qu'elle appelle une grammaire lexicale fonctionnelle. Ces chercheurs ont qualifié ce processus de *bootstrapping* sémantique. Dans la TSGN étendue, je fournis des arguments évolutionnistes, anatomiques et physiologiques explicites en faveur de l'idée selon laquelle les petits enfants possèdent déjà des catégories conceptuelles qui ne sont *ni* définies *ni* créées par des moyens ou des critères



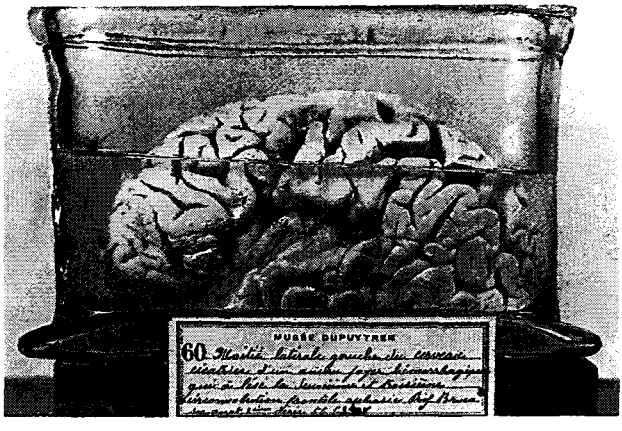


FIGURE 12-2

Aires du cerveau servant à la production du langage parlé (en haut). *Lorsque ces régions cérébrales sont affectées, il s'ensuit des aphasies qui peuvent prendre diverses formes. La photographie* (en bas) *montre le cerveau de l'un des patients de Paul Broca, qui avait une lésion au niveau de ce qu'on appelle l'aire de Broca. De son vivant, ce patient souffrait d'une aphasie motrice.*

sémantiques. Ces catégories sont nécessaires au déroulement du *bootstrapping* sémantique, et elles sont également nécessaires aux théories voisines, émises par Ronald Langacker, George Lakoff et d'autres, qui postulent l'existence de ce qu'on appelle des « grammaires cognitives » (voir la postface).

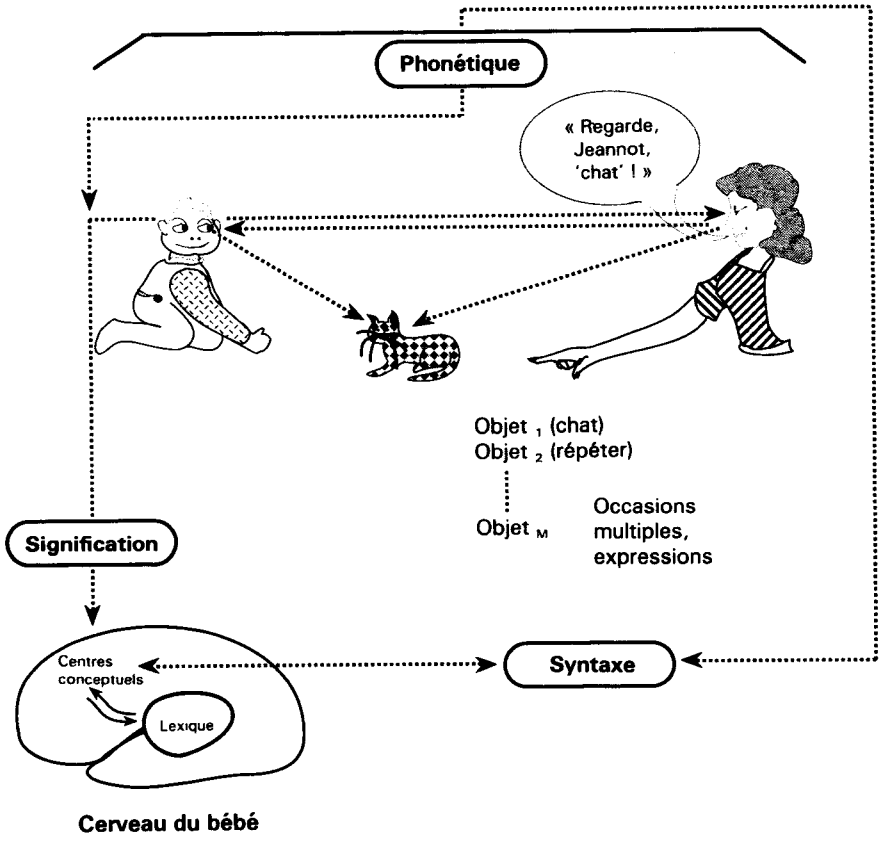


FIGURE 12-3

Bootstrapping sémantique. On voit ici, très schématiquement représenté, comment l'affect, les récompenses et l'apprentissage aboutissent à l'acquisition du langage sous l'effet des catégorisations. La phonétique fournit les moyens de relier les objets catégorisés aux significations. À mesure que des connexions réentrantes s'établissent avec les centres conceptuels, un bootstrapping sémantique a lieu. À mesure que se constitue un lexique et que des phrases sont rencontrées, la catégorisation de leur structure conduit à la constitution d'une syntaxe.

Ainsi, pour construire la syntaxe, c'est-à-dire les bases de la grammaire, le cerveau doit posséder des structures réentrantes qui permettent aux aspects sémantiques d'apparaître *en premier* (avant la syntaxe), à la faveur de la liaison entre symboles phonétiques et

concepts. Grâce au type de mémoire particulier fourni par les aires de Broca et de Wernicke, les niveaux phonétiques, sémantiques et syntaxiques peuvent interagir directement, mais aussi indirectement, via des circuits réentrants qui se forment entre ces aires du langage parlé et les aires cérébrales qui servent à la mémoire des valeurs-catégories. Lorsqu'un lexique suffisamment grand a été recueilli, les aires conceptuelles du cerveau classent par catégories l'*ordre* des éléments du langage, un ordre qui sera ensuite stabilisé dans la mémoire sous forme de syntaxe. En d'autres termes, le cerveau associe récursivement des séquences sémantiques à des séquences phonétiques puis engendre des correspondances syntaxiques, non pas à partir de règles préexistantes, mais en traitant les règles *en cours de développement dans la mémoire* comme des objets susceptibles d'être conceptuellement manipulés. La mémoire, la compréhension et la production du langage parlé interagissent de façons très diverses grâce à la réentrée. Cela permet la production de structures d'ordre supérieur (telles les phrases dans une grammaire) et contribue évidemment à l'élaboration de séquences d'ordre inférieur (telles les expressions). Bien sûr, une fois obtenue, la structuration séquentielle devient automatique, comme beaucoup d'autres actions motrices.

Contrairement aux humains, les chimpanzés ne possèdent pas les bases cérébrales nécessaires à la structuration séquentielle complexe de sons articulés. Ils semblent avoir des concepts et des pensées, et ils ont même des capacités « sémantiques » simples, mais comme il leur manque une syntaxe suffisamment élaborée, ils n'ont pas de langage intrinsèque.

La raison pour laquelle l'acquisition d'un véritable langage conduit à une énorme augmentation de la puissance conceptuelle est évidente. L'adjonction d'une mémoire symbolique spéciale, reliée à des centres conceptuels préexistants, donne lieu à la capacité d'élaborer, de raffiner, d'associer, de créer, et de se rappeler des quantités considérables de nouveaux concepts. Mais les centres du langage ne « contiennent » pas les concepts : ceux-ci ne « naissent » pas du langage. Au contraire, le sens naît de l'interaction de la mémoire des valeurs-catégories avec l'activité *conjointe* des aires conceptuelles et de celles du langage. Et, bien que le langage articulé ait probablement été nécessaire à la sélection des modifications de la morphologie du cerveau au cours de l'évolution, une fois celles-ci apparues, n'importe quel système gestuel (comme par exemple le langage des sourds-muets) aurait pu être utilisé si nécessaire par une

communauté linguistique donnée. De plus, comme beaucoup de systèmes cérébraux sujets à un développement épigénétique, le système sous-jacent à l'acquisition du langage n'est pas le même chez l'enfant et chez l'adulte. L'acquisition du langage a lieu à une période critique du développement. Et, selon toute vraisemblance, cette période a un rapport avec la sélection généralisée de synapses et de groupes neuronaux qui se déroule jusqu'à l'adolescence, car ensuite ce type de changements survient de façon différente et beaucoup moins étendue.

Dans la mesure où elle requiert l'évolution préalable de structures spécialisées, cette théorie du langage est une théorie nativiste. Mais elle ne fait appel à aucun nouveau principe hormis ceux de la TSGN. Ce n'est pas une théorie computationnelle, et elle ne prétend pas qu'il existe un dispositif d'acquisition du langage contenant des règles innées et génétiquement spécifiées pour la construction d'une grammaire universelle. La syntaxe se construit de façon épigénétique sous l'influence de contraintes génétiques, de même que les visages humains (qui sont à peu près aussi universels que la grammaire) se construisent sous l'influence de diverses contraintes imposées par le développement. D'ailleurs, les principes de la topobiologie (voir le chapitre 6) s'appliquent dans les deux cas.

Cette proposition est compatible avec la capacité de construire et d'interpréter un nombre potentiellement infini de phrases à partir d'un nombre fini de mots. En effet, le pouvoir de généralisation et de catégorisation d'un système conceptuel qui interagit de façon réentrante et récursive avec les aires spécialisées dans le langage est presque illimité. Et, dans la mesure où la syntaxe est construite à partir du sens (sous certaines contraintes), des relations grammaticales locales peuvent être construites même à partir de fragments de phrases qui ne sont pas soumis à l'ordre strict qui règne dans les phrases elles-mêmes. Une grammaire ainsi construite correspond nécessairement à l'activité continuelle d'un ensemble très précis de structures cérébrales, parmi lesquelles les plus importantes sont peut-être celles qui donnent naissance à la conscience primaire. Car en fait, si cette théorie est exacte, le langage ne peut exister sans la conscience primaire.

LA CONSCIENCE D'ORDRE SUPÉRIEUR

Munis de cette théorie du langage, nous pouvons revenir à notre sujet principal : la conscience d'ordre supérieur. Comment devient-

on « conscient d'être conscient » ? L'acquisition de cette capacité suppose que des systèmes de mémoire soient associés à la représentation conceptuelle d'un moi vrai (ou moi social) agissant sur un environnement et vice versa. Un modèle conceptuel de l'individualité doit être construit, ainsi qu'un modèle du passé. Et pour ce faire, il faut qu'un certain nombre d'étapes d'apprentissage modifient la relation de l'individu avec le présent immédiat au cours du développement.

En particulier, il faut des répertoires cérébraux capables de retarder les réponses. (On sait qu'il existe des répertoires de ce type dans le cortex frontal.) Ces répertoires doivent pouvoir classer par catégories les processus de la conscience primaire elle-même. Cela s'obtient principalement grâce à des moyens symboliques, par comparaison et récompense au cours de la transmission sociale et de l'apprentissage. Durant l'acquisition sémantique, cette récompense provient de l'association des symboles du discours à la satisfaction des besoins affectifs de l'individu par des congénères au cours des interactions avec ses parents, des activités de toilette et des activités sexuelles.

La figure 12-4 illustre la relation existant entre les aires du langage et les aires conceptuelles, qui permet le développement d'un concept de moi et de la conscience d'ordre supérieur ; cependant, il faut lui en adjoindre une autre (figure 12-3) illustrant les relations sociales. Le stockage à long terme des relations symboliques acquises au cours des interactions avec d'autres individus de la même espèce est essentiel au concept de moi. Cette acquisition s'accompagne d'une catégorisation des phrases associées au moi et au non-moi, et de leur liaison aux événements de la conscience primaire. L'élaboration correspondante issue de l'apprentissage d'éléments des mémoires phonémique et symbolique autorise également des catégorisations plus efficaces par l'intermédiaire des verbes de divers *actes* par rapport à soi et aux autres.

L'interaction de cet ensemble spécialisé de mémoires avec la mémoire conceptuelle des valeurs-catégories permet une modélisation du monde. Et, grâce à l'apparition de la faculté de faire la différence entre de tels modèles conceptuels-symboliques et l'expérience perceptive en cours, il devient possible de mettre au point un concept du passé. Cela libère l'individu de la tyrannie de l'instant présent et des événements survenant en temps réel. Le présent remémoré est intégré dans le cadre d'un passé et d'un futur.

Tandis que l'incarnation du sens et des références peut être associée à des objets et des événements réels par les connexions

réentrantes existant entre la mémoire des valeurs-catégories et la perception (conscience primaire), des interactions simultanées peuvent également avoir lieu entre une mémoire symbolique et les mêmes centres conceptuels. Une vie intérieure, fondée sur l'apparition de la parole au sein d'une communauté linguistique, devient possible. Elle est liée à des structures perceptives et conceptuelles, mais elle est hautement individuelle (elle est même personnelle), et fortement liée à l'affect et aux récompenses. En fait, il s'agit là de la conscience d'ordre supérieur, cette conscience capable de modéliser le passé, le présent, le futur, la personne et le monde.

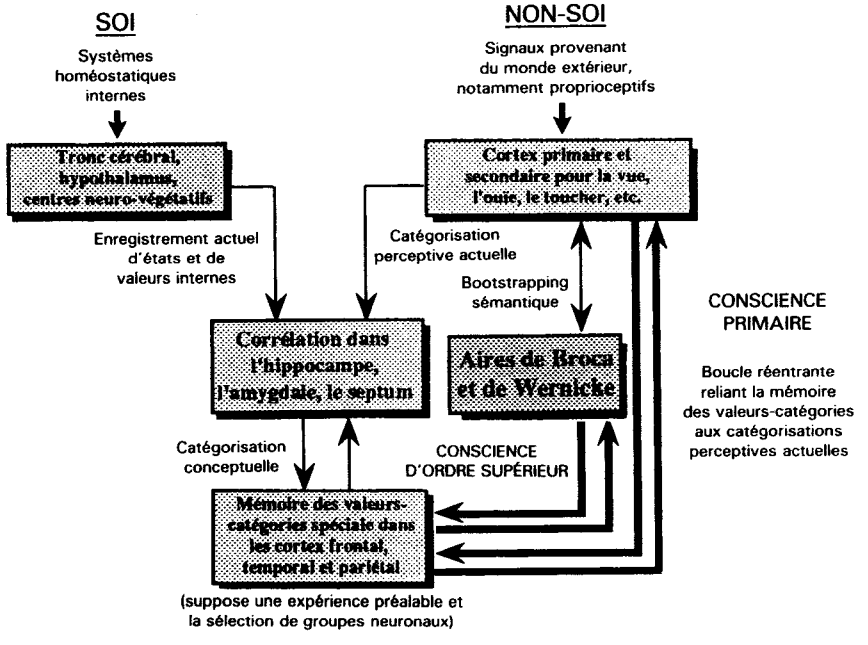


FIGURE 12-4

Comment apparaît la conscience d'ordre supérieur. (On peut joindre cette figure à la figure 11-1 sur la conscience primaire.) L'acquisition d'un nouveau type de mémoire via le bootstrapping sémantique (figure 12-3) entraîne une explosion conceptuelle. En conséquence, des concepts de soi, de passé et de futur peuvent être reliés à la conscience primaire. Il devient ainsi possible d'être « conscient d'être conscient ».

L'une des caractéristiques étonnantes de la conscience d'ordre supérieur est la rapidité avec laquelle elle est apparue. D'après les études paléontologiques, cela s'est en effet produit sur des intervalles très brefs à l'échelle de l'évolution. Les principes topobiologiques sous-jacents au développement du cerveau et les mécanismes de la TSGN permettent de rendre compte de cette rapidité, car ils auto-

risent d'énormes modifications de la taille du cerveau sur la période de temps relativement courte, à l'échelle de l'évolution, au cours de laquelle est apparu l'*Homo sapiens*. Selon la topobiologie, des modifications de synchronisation dans l'action des gènes morpho-régulateurs, à l'issue de mutations relativement peu nombreuses, suffisent à produire des modifications morphologiques relativement étendues (voir le chapitre 6). Les prémisses de la TSGN autorisent l'incorporation rapide de répertoires primaires nouveaux ou élargis dans les structures cérébrales existantes.

La figure 12-5 donne un aperçu sommaire de la relation entre conscience et évolution morphologique. Bien que très peu détaillé et dépourvu de repères temporels, ce diagramme suggère néanmoins comment *deux ensembles successifs de phénomènes de bootstrapping* (respectivement perceptif et sémantique), faisant chacun intervenir l'évolution de nouvelles morphologies (des circuits de mémoire et de nouvelles formes de réentrée), pourraient avoir donné lieu d'abord à la conscience primaire et ensuite à la conscience d'ordre supérieur.

Ce panorama évolutif suscite des questions supplémentaires concernant les avantages adaptatifs de la conscience. La conscience primaire fournit la capacité de déterminer, à partir de critères internes, l'importance de certains types de signaux parmi les multiples signaux parallèles issus d'environnements complexes. Cette importance est largement, mais pas entièrement, déterminée par l'histoire préalable et l'apprentissage de chaque animal. Quant à la conscience d'ordre supérieur, elle joint au tableau de cette identité biologique une individualité socialement construite. Le fait que certaines parties de la pensée consciente se libèrent des contraintes du présent immédiat et que la communication sociale s'enrichisse permet d'anticiper les états futurs et de planifier le comportement. Avec cette faculté viennent aussi la capacité de modéliser le monde, celles d'effectuer des comparaisons explicites et de peser les différentes issues possibles – et, à travers de telles comparaisons, de réorganiser des plans. Il est clair que ces capacités ont une valeur adaptative. En fait, l'histoire de l'humanité depuis l'apparition des chasseurs-cueilleurs reflète à la fois les propriétés adaptatives et non adaptatives de la seule espèce possédant une conscience d'ordre supérieur pleinement développée.

Certains anthropologues ont suggéré (de façon quelque peu fantasque) que si la taille de notre cerveau s'est accrue aussi rapidement, c'est parce que, à partir d'un certain point, la conscience d'ordre supérieur nous a conféré la capacité de nous duper nous-

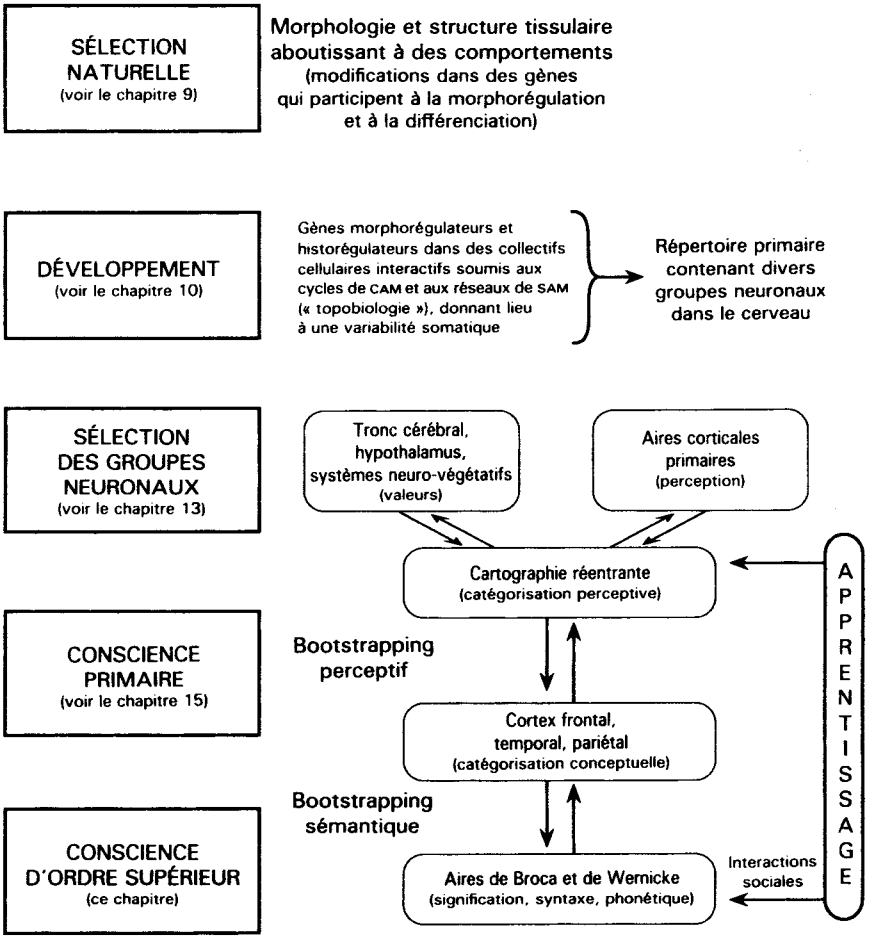


FIGURE 12-5

L'évolution de la conscience dépend de l'apparition de nouvelles morphologies. On voit ici une séquence d'événements survenant au cours de l'évolution, dans laquelle les principes de la sélection naturelle et du développement conduisent à des systèmes neuronaux de reconnaissance et donnent lieu à l'expérience consciente. Aucun nouveau principe, hormis ceux de la théorie de la sélection des groupes neuronaux, n'est nécessaire pour en rendre compte. Ce qu'il faut, en revanche, ce sont de nouvelles structures anatomiques *sélectionnées pour leurs fonctions à l'issue de l'évolution. Il s'agit notamment des structures qui apparaissent sur les deux premières figures de ce chapitre. Les principes sous-jacents à la fonction de chaque aire sont expliqués dans les chapitres indiqués sur la présente figure. Remarquez que la conscience primaire résulte d'un « bootstrapping perceptif », alors que la conscience d'ordre supérieur résulte d'un « bootstrapping sémantique ». Les deux bootstrappings reposent sur l'apparition de voies réentrantes adéquates dans le cerveau au cours de l'évolution.*

mêmes, de manière à duper les autres avec plus de « sincérité » et d'en profiter. Chez les animaux sociaux, cette duperie pourrait, selon ces auteurs, présenter des avantages adaptatifs. Une vieille histoire raconte que Boris et Ivan étaient à la gare. Ivan demanda à Boris : « Boris, où est-ce que tu vas ? » Et Boris lui répondit : « À Minsk. » Alors Ivan lui dit : « Boris, je te connais. Si tu allais vraiment à Minsk, tu me dirais que tu vas à Pinsk. Mais il se trouve que je sais que tu vas à Minsk. Alors pourquoi est-ce que tu me mens ? »

Munis de cette vision de l'animal humain, doué simultanément et interactivement d'une conscience primaire et d'une conscience d'ordre supérieur, nous pouvons à présent revenir à la fâcheuse question des sensations. Je rappelle que notre analyse théorique de la conscience était fondée sur trois hypothèses : l'hypothèse physique, l'hypothèse évolutionniste et l'hypothèse des sensations. Mais, puisque nous avons déjà admis l'idée que les êtres humains éprouvent des sensations, pourquoi revenir à cette question ? Nous savons en effet qu'il est impossible d'adopter un point de vue extérieur – c'est-à-dire de supposer que la théorie permette, à travers une description, à un animal imaginaire dépourvu de sensations de savoir ce qu'est une sensation. Nous en avons également dit assez sur les mécanismes de la conscience pour savoir que seul le fait de posséder la morphologie et l'expérience adéquates permet à un individu d'éprouver des sensations. Cependant, il se trouve que notre vision élaborée des choses nous fournit un certain nombre de précisions supplémentaires.

Tout d'abord, elle nous permet de voir clairement comment les *différentes* sensations sont *discriminées* : à travers les différences de structure neurale et de comportement des différentes voies sensorielles. On sait cela depuis longtemps – en fait, depuis la doctrine des énergies nerveuses spécifiques énoncée par Johannes Müller. Nous pouvons ajouter qu'un animal doté d'une conscience d'ordre supérieur est susceptible d'attribuer à deux états phénoménaux donnés des noms différents dès lors que ces états auront été évoqués par des voies neuronales différentes.

Si les animaux qui ne possèdent qu'une conscience primaire éprouvent aussi des sensations, ils sont incapables de les décrire explicitement, que ce soit à un observateur humain ou à eux-mêmes, puisqu'il leur manque un moi conceptuel. Comme des flashes éclairant une pièce, leurs sensations, si tant est qu'elles existent, ne persistent que pendant la durée du présent remémoré de la scène.

Et nous ne pouvons invoquer comme preuve de leur éventuelle existence que leurs réactions comportementales.

Mais en ce qui nous concerne, les choses se passent autrement. Nos sensations, qui nous sont personnelles, sont des recatégorisations, effectuées par la conscience d'ordre supérieur, de relations perceptives chargées de valeurs et établies au sein de chaque modalité sensorielle, ou de leurs combinaisons conceptuelles. Nous les décrivons aux autres de façon imparfaite ; à nous-mêmes, nous pouvons les décrire de façon plus directe. Cet ensemble de relations est habituellement, mais pas toujours, relié à des valeurs. *Le fait de s'être libéré des contraintes temporelles autorise la localisation dans le temps* d'états phénoménaux par un moi qui souffre ou se réjouit. Et la présence d'un langage adéquat améliore énormément la discrimination ; le talent des taste-vin, par exemple, peut être considéré comme le résultat d'une passion fondée sur des sensations qui deviennent de plus en plus fines grâce au langage.

Cette vision de la conscience d'ordre supérieur nous permet d'entrevoir ce qu'il y a sous le moi qui relie la phonétique à la signification au sein d'une phrase dénominative. Une fois le moi mis en place à l'issue des interactions sociales et linguistiques, avec la conscience primaire pour base, il se crée un monde qui *requiert* des noms et des intentions. Ce monde reflète les événements intérieurs que l'on se rappelle, les événements imaginaires, et aussi les événements extérieurs que l'on éprouve perceptivement. La tragédie devient possible – la perte du moi due à la mort ou à la maladie mentale, la mémoire de douleurs inapaisables. Et de ce fait même, les grands actes de création et une imagination inépuisable prennent forme.

Ironiquement, le moi est ce qui est le moins bien compris par son détenteur, même muni d'une théorie de la conscience. Étant donné la manière dont la conscience d'ordre supérieur apparaît et celle dont les dénominations se déroulent, ce n'est pas étonnant – si ce n'est pour chacun de nous en tant que détenteurs de ce type de conscience. L'incarnation impose inévitablement des limites. Le désir de dépasser ces limites crée des contradictions, des illusions, et une mystique qui fait de l'étude de l'esprit un défi majeur, car à partir d'un certain point, du moins en ce qui concerne ses créations individuelles, l'esprit se trouve hors de portée de la science. La recherche scientifique reconnaît cette limitation et ne se laisse

pas aller à des exercices mystiques ou à des illusions. Après tout, cette limitation existe pour une raison bien simple : les formes d'incarnation qui mènent à la conscience sont propres à chaque individu, à son corps et à son histoire individuelle.

Chapitre 13

L'attention et l'inconscient

> *Les illusions s'imposent à nous parce qu'en nous permettant d'éprouver du plaisir elles nous épargnent bien des peines. Par conséquent, nous devons accepter sans nous en plaindre qu'elles soient réduites en miettes lorsque parfois elles se heurtent à un peu de réalité.*
>
> Sigmund Freud

> *La conscience règne et ne gouverne pas.*
>
> Paul Valéry

Ce que j'ai dit jusqu'ici souffre deux objections. Tout d'abord, je n'ai pas *vraiment* expliqué « ce que l'on ressent » quand on est conscient. Ensuite, il y a beaucoup de choses que je n'ai pas expliquées, comme par exemple le fait qu'une bonne part de notre comportement soit régie inconsciemment. Sigmund Freud a consacré une grande partie de sa vie à essayer de comprendre ce phénomène, et notamment le refoulement des expériences menaçantes, douloureuses ou désagréables. Je vais tout d'abord répondre brièvement à la première objection, puis je m'attarderai davantage sur la seconde.

Étant donné que la conscience nous est essentielle pour savoir que nous existons – comme l'affirmait Descartes –, il n'est pas surprenant que nous attendions beaucoup de toute description qui prétend l'expliquer. De mon explication, on pourrait donc dire : « Vous croyez avoir expliqué comment la mémoire, la catégorisation perceptive, la réentrée, etc., donnent naissance aux *propriétés* de la conscience, mais vous n'avez pas expliqué ce que *je* ressens en tant qu'être conscient, ou encore pourquoi *je* me sens conscient. La conscience est étrange, mystérieuse, elle constitue l'ultime mystère. » Pour répondre à cette objection, je dois préciser les limites de toute tentative d'explication scientifique, et ensuite montrer ce qu'une explication de la conscience a nécessairement de particulier.

La science s'intéresse aux corrélations formelles entre diverses propriétés, et s'attache à développer des édifices théoriques capables de décrire tous les aspects connus de ces corrélations, sans exception, de la manière la plus parcimonieuse et la plus utile possible. Elle

doit énoncer ses descriptions de sorte que deux observateurs humains, quels qu'ils soient, puissent se les communiquer et les comprendre. Toute description dont la cible n'est pas l'observateur humain, conscient et capable de comprendre, susceptible de détecter des failles logiques, de reproduire des expériences et d'en construire d'autres, n'est pas une description scientifique. Un exemple de description non scientifique consisterait de ma part à raconter mes sensations personnelles, mes souvenirs apparents et mes émotions durant une transe induite par des drogues. Au mieux, un observateur pourrait tenter de corréler les descriptions fournies par une vingtaine de sujets en transe (dont moi-même) et y détecter des régularités. Mais cet observateur serait incapable de corréler de façon fiable et détaillée *mes* sentiments particuliers, mon histoire et ma façon particulière d'oublier, ou de les généraliser avec certitude. Par conséquent, même s'il peut arriver que la science parvienne à discerner les éléments communs à vingt récits différents, elle est incapable de décrire des histoires individuelles.

La conscience est un sujet scientifique tout à fait particulier, car elle est elle-même le processus individuel, personnel, qui doit fonctionner en chacun de nous pour qu'il puisse procéder à la *moindre* explication scientifique. Même si je ne me rends pas compte de ce que j'ai pu oublier ou refouler, même si je ne connais pas les facteurs inconscients qui gouvernent mon comportement, j'ai l'impression que le processus de la conscience est d'un seul tenant – du moins lorsque je suis dans mon état normal. Il est donc naturel que j'exige une explication de ma propre conscience en des termes qui me soient personnellement satisfaisants. Mais je dois me rendre compte qu'il ne s'agit pas là d'un acte scientifique, et d'ailleurs je ne m'attends pas à ce qu'il le soit. Après tout, personne ne dit à un physicien : « Vous avez expliqué l'énergie et la matière en termes de relations de symétrie, et vous êtes presque remonté, dans vos théories, jusqu'à l'origine de l'univers. Mais vous n'avez pas *vraiment* expliqué *pourquoi* il existe des choses, *pourquoi* il n'y a pas que du vide. » En fait, il serait vain de tenter de fournir une telle explication, et étant donné les circonstances, aucune science expérimentale ne pourrait prétendre y parvenir mieux qu'une autre. Il est impossible, dans ce cas, de fournir une explication scientifique. Mais alors, pourquoi sommes-nous tentés d'exiger une explication scientifique de ce que ressent personnellement un être conscient ? Parce que nous avons des certitudes personnelles sur notre conscience et sur ses liens avec la notion de moi. Cela nous donne envie

d'être plus exigeants avec les psychologues qu'avec les physiciens ou les astrophysiciens. Mais nos exigences ne sont pas raisonnables du point de vue scientifique.

La réponse à la question serait formellement la même que celle du physicien : « Je vous ai donné une théorie fondée sur des structures et des relations connues, sur des données expérimentales. Elle stipule que, si vous effectuez une opération sur des structures considérées comme importantes pour les propriétés de la conscience, celles-ci seront modifiées d'une façon prévisible et pourront même disparaître. Si, par exemple, vous coupez une boucle réentrante desservant une partie du cerveau essentielle à la reconnaissance des visages (et seulement cette partie) chez un individu conscient, l'individu souffrira de prosopagnosie. Il sera désormais *inconscient* du fait qu'il est encore capable (comme le montre l'examen de la mémoire implicite) de reconnaître des visages dont l'observateur sait que l'individu les a vus auparavant, et il niera tout à fait sincèrement les avoir reconnus. Et il en va de même avec d'autres tests ; ainsi, une théorie de la conscience *peut* avoir des composants opérationnels. Mais, pour être une bonne théorie, elle doit également parvenir à unifier toutes sortes de faits pertinents et nous permettre de mieux comprendre la conscience. (Par exemple, elle doit permettre d'expliquer comment il se fait que certains processus puissent être inconscients et motiver quand même des comportements – ce qui est le but de ce chapitre.)

Pourquoi tenons-nous donc tant à adopter un point de vue extérieur malgré toutes ces explications ? Pourquoi existe-t-il une mystique de la conscience – un désir d'explication universelle, un besoin de préserver indéfiniment à la conscience son caractère d'expérience individuelle ? Une réponse raisonnable à cette question consiste à dire que chaque conscience dépend de son histoire et de son incarnation, lesquelles sont particulières et uniques. Le moi conscient humain se construit – ce qui est assez paradoxal – à travers les interactions sociales, bien qu'il ait été sélectionné, au cours de l'évolution, principalement pour atteindre ses objectifs et satisfaire ses exigences personnelles. Dès lors, en tant qu'individus, nous souhaitons avoir une explication que la science ne peut nous fournir. De même, nous aspirons à l'immortalité. Mais notre incapacité en tant que scientifiques à expliquer chaque conscience individuelle n'est pas plus mystérieuse que notre incapacité à expliquer pourquoi il existe quelque chose plutôt que rien. Ou plutôt, il y a peut-être là un mystère, mais il n'est pas d'ordre scientifique. En effet, si

l'on ne considère que son propre esprit, le mystère réside alors dans le fait qu'on a du mal à imaginer comment cet esprit-là a pu apparaître relativement à sa propre histoire personnelle. Nous sommes « enfermés à l'intérieur ».

Il existe néanmoins un moyen, difficile à mettre en pratique mais bien réel, d'aborder scientifiquement cette clôture de la conscience individuelle, qui est à la source du « mystère ». Si l'on parvenait à construire un objet dont les structures et l'expérience lui permettraient à la fois de devenir conscient et d'avoir un langage, on pourrait déterminer s'il éprouve ou non des sensations. Mais si cet objet venait à décrire un sentiment donné, serait-il raisonnable (et éthique) de le reconstruire sans y intégrer les structures censées être essentielles à ce sentiment ? Et, à l'issue d'une telle procédure, étant donné que son moi aurait été construit par interaction de processus inconscients et d'interactions « sociales » conscientes, le même objet se sentirait-il « bizarre » et « différent » ? Seul l'avenir le dira, mais aussi fantastique et improbable que cette idée puisse paraître, elle est du moins théoriquement envisageable. En revanche, on ne peut pas en dire autant de la possibilité de réaliser des expériences sur la création de l'univers.

Avant d'aborder les processus inconscients qui donnent naissance à la conscience et qui la modifient, il me semble utile de distinguer ce que l'on peut facilement imaginer à propos de l'esprit de ce que l'on ne peut pas. Comme nous l'avons vu, toute tentative pour expliquer l'esprit se heurte à une difficulté majeure : l'esprit est le produit d'interactions physiques se déroulant à des niveaux d'organisation extrêmement nombreux, qui vont du moléculaire au social. De plus, ces interactions relèvent souvent d'idiosyncrasies irréversibles ; les caractéristiques structurelles essentielles à leurs mécanismes comportent des cartographies parallèles, soit entre une carte et plusieurs autres, soit entre plusieurs cartes et plusieurs autres. Or nos cerveaux (et en particulier ceux des philosophes) ne sont pas très doués pour visualiser des organisations aussi complexes. Néanmoins, la situation n'est peut-être pas désespérée ; je reviendrai plus loin sur la façon dont l'avènement d'ordinateurs de plus en plus puissants pourrait nous permettre de construire des heuristiques susceptibles de nous montrer comment les choses s'assemblent.

En attendant que ce domaine soit plus largement développé, nous pouvons nous demander ce qu'il est facile d'imaginer à propos de

l'esprit. Je pense que la plupart des gens seront d'accord avec la liste suivante :

1. Le fonctionnement grosso modo des circuits cérébraux en termes d'entrées et de sorties : c'est ce qu'étudie la neurophysiologie classique.

2. L'interaction entre les types de comportements de l'animal et le monde physique des stimulus : c'est ce qu'étudie la psychologie descriptive.

3. Certains actes de transmission sociale. Exemples : l'étude de l'*imprinting* [1] en éthologie et les propositions généralement admises de la « psychologie populaire » – ce que les gens semblent croire, désirer, leurs intentions.

Qu'est-ce qui est difficile à comprendre ?

1. Le résultat net de l'action simultanée et parallèle de populations neuronales complexes. La difficulté de prédire le résultat de l'activité d'un grand nombre de groupes neuronaux illustre bien le problème.

2. La mémoire en tant que processus dynamique et propriété systémique, c'est-à-dire qui n'est *pas* équivalente à la somme des modifications synaptiques qui lui sont sous-jacentes. Nous pouvons prendre comme exemple la réponse globale d'un automate comme Darwin III à la suite d'un apprentissage.

3. Des phénomènes psychologiques plus complexes telle la conscience. On trouvera de nombreux exemples dans les deux précédents chapitres.

4. L'idée d'un moi socialement constitué, résultant des interactions de processus à la fois conscients et inconscients. Nous allons consacrer la suite de ce chapitre à l'illustration de cette idée.

Il existe certainement d'autres listes possibles. Cependant, pour comprendre les processus psychologiques (et en particulier ceux de la seconde liste) dans le cadre d'une théorie du cerveau, on doit non seulement disposer d'une bonne théorie, mais aussi être capable d'utiliser des modèles informatiques synthétiques pour vérifier la cohérence interne de ses mécanismes et analyser des interactions à plusieurs niveaux. Malheureusement, les mathématiques à elles seules ne contribuent pas autant à nos constructions verbales qu'à

---

1. *Imprinting* : forme d'apprentissage qui a lieu rapidement, à un certain moment du développement, et dont les effets au niveau du comportement sont extrêmement durables. L'exemple classique est celui des canards nouveau-nés qui suivront comme s'il s'agissait de leur mère un objet qui leur aura été présenté pendant la période critique (N.d.T.).

notre physique. Actuellement, la voie la plus prometteuse est en fait celle des simulations par ordinateur.

Avant d'aborder ce qui suit, je suggère fortement au lecteur de déployer tous les efforts nécessaires pour bien comprendre comment les couples de classification, associés à la réentrée, donnent lieu à la catégorisation perceptive, et aussi comment la mémoire dynamique fonctionne en tant que propriété systémique (voir les chapitres 9 et 10). Une fois ces deux processus bien compris, je pense que le lecteur pourra parcourir pas à pas les modèles que j'ai décrits concernant la conscience primaire et la conscience d'ordre supérieur, et qu'il « verra » alors comment elles pourraient être susceptibles de donner lieu à des propriétés jamesiennes, de modifier l'importance relative des choses dans une scène et de permettre la formulation de plans. La meilleure façon de s'y prendre consiste peut-être à réexaminer les figures des chapitres correspondants. La compréhension ainsi acquise permettra de voir beaucoup plus aisément comment opèrent l'attention et l'inconscient.

L'ATTENTION

Dans ma description de la conscience, je n'ai pas abordé explicitement la question de l'attention, dont James disait qu'elle était « la prise de possession par l'esprit, sous une forme claire et vive, d'un objet ou d'une suite de pensées parmi plusieurs qui semblent simultanément possibles ». L'attention et la conscience ne sont pas la même chose, mais le lien entre elles pose quelques-uns des problèmes théoriques les plus ardus qui soient. Par exemple, il faut distinguer l'attention de l'état de veille, car il ne s'agit pas simplement d'une question de vigilance ou d'éveil : l'attention confère une composante *directionnelle* au comportement, et elle module la réactivité de l'animal à son environnement. Plus encore, l'attention met en évidence la « fragilité » de la conscience : elle concentre l'esprit sur ses objets et abolit, ou du moins atténue, le « superflu » environnant. En effet, il ne semble possible de vouer une attention spécifique qu'à quelques objets ou suites de pensées en même temps. L'attention est hautement sélective : il semble même que ce soit là une des caractéristiques nécessaires.

De nombreuses théories de l'attention sélective sont fondées sur la notion de « filtrage » des signaux d'entrée, soit précoce, soit tardif. Pourtant, divers résultats donnent à penser qu'un tel filtrage n'existe

pas. J'ai moi-même tendance à croire, comme certains l'ont dit, que les mécanismes cérébraux de l'attention sont apparus à la faveur de la pression évolutive qui s'est exercée sur les animaux et qui les a poussés à sélectionner une action au sein d'un ensemble d'actions adéquates. Un animal qui a faim ou qui se sent menacé doit sélectionner un objet ou une action parmi de nombreux objets ou actions possibles. Il est clair que la capacité de choisir rapidement un mode d'action, et de s'y tenir à l'exclusion de tous les autres, confère un avantage sélectif considérable. Le fait d'en être capable permet d'atteindre un objectif sans être perturbé par d'éventuelles tentatives pour engager simultanément deux actions incompatibles. Cette capacité peut donc être essentielle à la survie de l'animal.

Cette théorie « motrice » de l'origine de l'attention n'implique pas que les composantes perceptives ne soient pas importantes. En fait, il est évident qu'il existe de multiples mécanismes d'attention, qui vont de la concurrence perceptive au choix volitif. Mais si le résultat final consiste en la formulation d'une séquence d'actions ou de plans moteurs – que ceux-ci soient ou non exécutés –, alors, d'après la TSGN, il est probable que les cartographies globales et les ganglions de la base interviennent (voir les chapitres 9 et 10, et en particulier la figure 10-2). En effet, chez l'animal doté d'une conscience primaire, un équilibre doit être atteint entre le fait de réagir à des choses dont l'importance a été décidée au niveau interne et le fait de répondre aux nouveautés survenant à l'extérieur. Et, en présence d'une conscience d'ordre supérieur, la situation devient beaucoup plus riche. Les états volitifs associés à la sélection de plans, de valeurs et de projections temporelles sont tous susceptibles de modifier les contributions relatives des différentes parties d'une cartographie globale. Quoi qu'il en soit, dans les deux cas, il est probable que de larges portions du système nerveux interviennent au moment où il s'agit de modifier une cartographie globale afin de déplacer l'attention d'une chose vers une autre.

Selon ce point de vue, nous devrions nous attendre à ce que l'attention se modifie à la suite de changements survenant à différents niveaux d'une cartographie globale – à la suite d'activités conscientes autant qu'inconscientes. Mais comment un tel système fonctionnerait-il au juste ? Tout modèle destiné à expliquer le phénomène d'attention doit rendre compte de sa sélectivité ; du fait qu'une fois qu'un animal a acquis une compétence, celle-ci devient automatique ; de la capacité d'interrompre des actions automatiques

pour prêter attention à des nouveautés ; et enfin, de la capacité d'orienter spécifiquement l'attention par des moyens conscients.

Nous savons tous que l'attention consciente joue un rôle important dans l'acquisition de compétences élaborées. Mais, dans de nombreux cas, un apprentissage réussi nous permet ensuite d'effectuer des actions complexes sans y prêter attention. Nos réalisations demeurent alors inconscientes tant qu'aucune nouveauté ou menace ne vient nous solliciter davantage. Souvenez-vous que, lorsque j'ai présenté la TSGN, j'ai suggéré que les ganglions de la base étaient parmi les principaux organes de la succession, agissant avec le cortex pour choisir des plans moteurs. Les plans moteurs, qui peuvent être consciemment formés chez l'être humain, sont exécutés via le cortex moteur à mesure que celui-ci envoie des signaux à la moelle épinière. Mais les sorties du cortex sont également acheminées vers les ganglions de la base. Et si en revanche, dans l'autre sens, ces structures ne sont qu'indirectement connectées au cortex, cette connexion est extrêmement importante. En effet, les sorties des ganglions de la base sont inhibitrices : elles peuvent donc également *inhiber l'inhibition.* En d'autres termes, ces sorties peuvent *désinhiber* des régions cibles du cortex, ce qui les excite, ou les prépare à recevoir une entrée excitatrice, état important du point de vue de l'attention.

En accord avec un plan donné, les ganglions de la base désinhibent sélectivement les noyaux thalamiques qui envoient leurs projections vers le cortex. Cela conduit à une excitation anticipée et sélective des aires corticales correspondant au programme moteur. Ces aires corticales deviennent alors plus sensibles aux entrées sensorielles qui sont compatibles avec la réalisation de la tâche via une cartographie globale. Un tel mécanisme peut rendre compte du phénomène de concentration de l'attention.

Que dire du fait qu'une activité automatique puisse être interrompue par des nouveautés ? Si la tâche n'est pas achevée au bout d'un certain temps, ou si un événement inédit est détecté et une catégorisation effectuée, des signaux d'« alarme » pourront être transmis aux systèmes de valeurs du mésencéphale, qui renvoient des signaux vers le cortex et les ganglions de la base. Ces systèmes pourront alors renvoyer des signaux afin d'interrompre le plan moteur dans le cortex et bloquer l'exécution d'un programme moteur. Tant qu'une action automatisée est accomplie sans le moindre accroc, ces noyaux du mésencéphale n'interviennent pas. Mais, dans le cas contraire, à l'instar de quelqu'un qui, au cours d'une conversation

avec le chauffeur d'une voiture, lui crie soudain « Attention ! », ils provoqueront un déplacement de l'attention.

Mais comment la *conscience* parvient-elle à modifier l'attention et l'ordre de priorité dans la construction des cartographies globales ? Dans le cas de la conscience primaire, elle peut le faire en modifiant la proéminence des différentes boucles réentrantes parallèles reliées aux ganglions de la base, grâce à un processus semblable à celui que nous venons de décrire. Dans le cas de la conscience d'ordre supérieur, des schémas verbaux existant dans les aires conceptuelles peuvent, par l'intermédiaire des activités du cortex frontal et du système limbique, contrôler l'allocation de la désinhibition opérée par les ganglions de la base, qui sont fortement connectés à ces régions.

La fragilité de l'attention est une question particulièrement intéressante. Comment se fait-il que l'attention consciente soit si étroite qu'elle ne soit, en général, capable de se concentrer que sur une, tout au plus deux cibles en même temps ? La théorie motrice, qui considère l'attention comme un produit des besoins évolutifs, nous suggère une réponse possible. D'une part, en effet, les plans et les programmes moteurs sont plus ou moins exclusifs (c'est-à-dire, ils n'admettent pas d'actions contradictoires simultanées). D'autre part, étant donné les grandes quantités de tissu nerveux intervenant dans chaque cartographie globale, il semble probable que l'on ne puisse maintenir simultanément qu'un petit nombre de cartographies complexes sans qu'elles interfèrent les unes avec les autres.

Une telle vision de l'attention confère tout de même une importance primordiale aux mécanismes non conscients et aux comportements d'orientation permis par les cartographies globales en réponse à des situations d'urgence. Mais puisque le fait d'avoir des états conscients intentionnels dépend de l'existence de valeurs, de catégories et de souvenirs autant que de plans, cette vision sélectionniste de l'attention nous rend également capables d'avoir consciemment l'« intention de faire attention » à ce que nous avons planifié ou envisagé. Cependant, cette capacité est toujours susceptible d'entrer en concurrence avec des éléments inconscients et non conscients (les derniers étant ceux qui ne deviennent *jamais* conscients). Nous savons tous que nous pouvons faire des lapsus et agir de façon « imprévue ». Ces deux exemples suggèrent bien que des processus inconscients interviennent dans le phénomène d'attention.

L'INCONSCIENT

C'est Freud qui a surtout contribué à mettre en évidence l'importance des processus inconscients dans nos comportements et nos sentiments. Dans son *Esquisse de psychologie scientifique*, il a tenté de fournir une description neuronale explicite de la relation existant entre le comportement et les processus conscients et inconscients, mais a très vite renoncé. Par la suite, il a surtout



FIGURE 13-1

Sigmund Freud (1856-1939), fondateur de la psychanalyse et explorateur des mécanismes de refoulement au niveau de la mémoire.

donné des explications psychologiques des comportements qui mettaient l'accent sur l'intentionnalité tout en étant impitoyablement déterministes.

Le postulat de l'existence d'un inconscient est un principe unificateur des théories psychologiques de Freud. Depuis son époque, de nombreux résultats, issus de l'étude des névroses, de l'hypnotisme et des lapsus, se sont accumulés qui montrent que les thèses freudiennes de base sur l'action de l'inconscient étaient, pour l'essentiel, correctes. Dans le sens utilisé par Freud, le mot inconscient faisait référence aussi bien à des éléments qui peuvent aisément être transformés en états conscients – le « pré-conscient » – qu'à ceux qui ne peuvent l'être qu'au prix de grandes difficultés, ou pas du tout – « l'inconscient proprement dit ». Freud a suggéré que des événements menaçants pouvaient être refoulés dans la mémoire, de sorte qu'ils ne puissent être rappelés à la conscience.

Nous ne devons pas oublier qu'il s'agit là de termes psychologiques et non structurels. Mon défunt ami Jacques Monod – grand biologiste moléculaire – et moi-même avions souvent des discussions animées à propos de Freud. Il soutenait mordicus que Freud était anti-scientifique et que c'était probablement un charlatan. De mon côté, je défendais l'idée que, tout en n'étant peut-être pas scientifique dans notre sens du terme, Freud a été un grand pionnier intellectuel, en particulier en ce qui concerne sa vision de l'inconscient et de son rôle dans le comportement. Monod, issu d'une austère famille protestante, répondait à cela : « Je suis tout à fait conscient de mes motivations et entièrement responsable de mes actes. Ils sont tous conscients. » Exaspéré, je lui rétorquai un jour : « Jacques, disons tout simplement que tout ce que Freud a dit s'applique à moi et que rien ne s'applique à toi. » « Exactement, mon cher ami », répondit-il.

La notion freudienne de refoulement est compatible avec les modèles de la conscience présentés ici. En effet, la TSGN étendue fait fortement intervenir des systèmes dépendant des valeurs dans la formation de la mémoire. En outre, la discrimination moi/non-moi (voir la figure 11-1) requiert la participation de systèmes de mémoire qui sont à tout jamais inaccessibles à la conscience. Ainsi, le refoulement – l'incapacité sélective de se souvenir – serait sujet à des recatégorisations fortement chargées de valeurs. Étant donné que la conscience d'ordre supérieur se construit à travers les interactions sociales, il serait avantageux, du point vue de l'évolution, d'avoir des mécanismes capables de refouler les recatégorisations

qui mettent en péril l'efficacité des concepts du moi. On sait que des circuits qui interagissent avec les systèmes de valeurs existent dans l'hippocampe et dans les ganglions de la base. Chez l'animal doué de parole, les symboles sont importants, et l'apparition au cours de l'évolution d'une façon de réduire l'accès à des états considérés comme menaçants pour le concept de moi aurait une valeur sélective. Cela fournit un indice fondamental sur les propriétés des émotions, un thème que nous effleurerons dans un chapitre ultérieur.

Ma conclusion générale – importante pour toutes les théories de l'esprit – est la suivante : étant donné qu'il existe des actes régis par l'inconscient, les conclusions auxquelles on parvient par introspection consciente peuvent être sujettes à de graves erreurs. Autrement dit, le cartésianisme invétéré est incompatible avec les faits. Descartes, un adulte génial qui dominait le langage, n'a pas tenu compte d'un certain nombre d'événements. Le premier est que la nature de la conscience d'ordre supérieur est déterminée par le développement. Rappelez-vous qu'il est peu probable qu'un bébé français, aussi doué soit-il, s'écrie « Je pense, donc je suis ». Le deuxième est que la notion cartésienne de conscience fondée sur le langage *n'est pas* auto-suffisante et incontestable. Puisqu'elle est linguistique, elle entretient un dialogue permanent avec un « autre », même si cet interlocuteur n'est pas présent. Le troisième est que des mécanismes inconscients bloquent et perturbent ce que nous considérons comme des suites de pensées transparentes et évidentes. En posant les questions qui suscitèrent son doute méthodique, et en exposant courageusement ses pensées sur la nature de l'esprit, Descartes est devenu l'un des grands pionniers des recherches modernes en philosophie et en psychologie. Depuis la présentation de sa méthode, cependant, des connaissances se sont accumulées qui nous obligent à adopter une position beaucoup plus humble à propos de la certitude de ce que nous savons.

Il me semble opportun de redire ici à quel point nous savons peu de choses. Cela ne doit pas nous surprendre, puisque nous avons vu à quel point il est difficile d'imaginer l'esprit et la stratification complexe de ses mécanismes et de ses processus. Mais nous avons à présent à notre disposition un grand nombre de ressources qui manquaient à Descartes. Celles-ci devraient, au bout du compte, nous permettre d'apprécier à quel point sa théorie, bien que « fausse », aura été fructueuse, puisqu'elle nous aura poussés à tenter de comprendre l'esprit.

Chapitre 14

Les niveaux et les boucles : un résumé

> *J'ai l'impression que la race humaine se trouve à la veille d'une percée majeure. Nous sommes parvenus au point où nous devenons capables de lever un coin du voile qui nous masque la nature de notre propre esprit. Est-il encore concevable que nous retirions la main et fassions marche arrière par découragement et manque d'imagination ?*
>
> Percy Williams Bridgman

Il est grand temps d'envisager une autre façon de voir les phénomènes mentaux, de bâtir un modèle neuroscientifique de l'esprit. Si celui que je propose ici est nouveau, c'est parce qu'il se fonde sans aucun remords sur la physique et la biologie. Il s'appuie également sur les notions d'évolution morphologique et de sélection, et récuse l'idée selon laquelle une description syntaxique des opérations et des représentations mentales (voir la postface) suffirait à expliquer l'esprit. D'autres que moi ont tenu des propos semblables, mais ils ne les ont pas intégrés dans une théorie unique fondée sur l'évolution, une théorie reliant l'embryologie, la morphologie, la physiologie et la psychologie. Or seule une telle théorie de l'esprit est susceptible d'être réfutée par des moyens scientifiques.

Le chemin qui joint ces diverses disciplines est accidenté et, comme le lecteur a pu le constater, il est parfois difficile à parcourir. Cela est dû au fait que la construction de l'esprit requiert un très grand nombre de niveaux d'organisation différents, ainsi qu'un très grand nombre de boucles interactives servant à relier des niveaux de description qui à première vue peuvent sembler disparates. Mais, puisque l'esprit résulte d'une évolution et non d'une planification logique, je ne pense pas que les choses puissent se passer autrement. C'est cette profusion de niveaux, et non pas un quelconque nouveau principe ésotérique issu de la physique ou de la théosophie, qui fait qu'il est difficile de réfléchir sur l'esprit. Le cerveau, qui donne naissance à l'esprit, est le prototype d'un système complexe et, du point de vue de son style d'organisation, il ressemble davantage à une jungle qu'à un ordinateur. Cependant, cette analogie comporte

une faille. En effet, alors que dans une jungle les plantes sont sélectionnées au cours de l'évolution, la jungle elle-même ne l'est pas. En revanche, le cerveau est soumis à deux processus de sélection – à la sélection naturelle et à la sélection somatique.

Il en résulte une entité subtile et multistratifiée, pleine de boucles et de niveaux différents. Des gènes aux protéines, des cellules au développement orchestré, de l'activité électrique à la libération de neuromédiateurs, des couches sensorielles aux cartes, de la forme à la fonction et au comportement, et, en sens inverse, de la communication sociale vers n'importe lequel de ces niveaux ou vers tous à la fois, nous nous trouvons devant un système de sélection somatique qui est constamment soumis à la sélection naturelle (figure 14-1). Dans ces conditions, il n'est pas étonnant que les philosophes, qui réfléchissent au problème de l'esprit sans disposer de ces connaissances, aient été tentés de postuler l'existence de certaines entités, que les physiciens soient tentés de postuler l'existence de nouveaux champs matériels exotiques, ou que ceux qui aspirent à l'immortalité continuent à postuler l'existence d'esprits éternels.

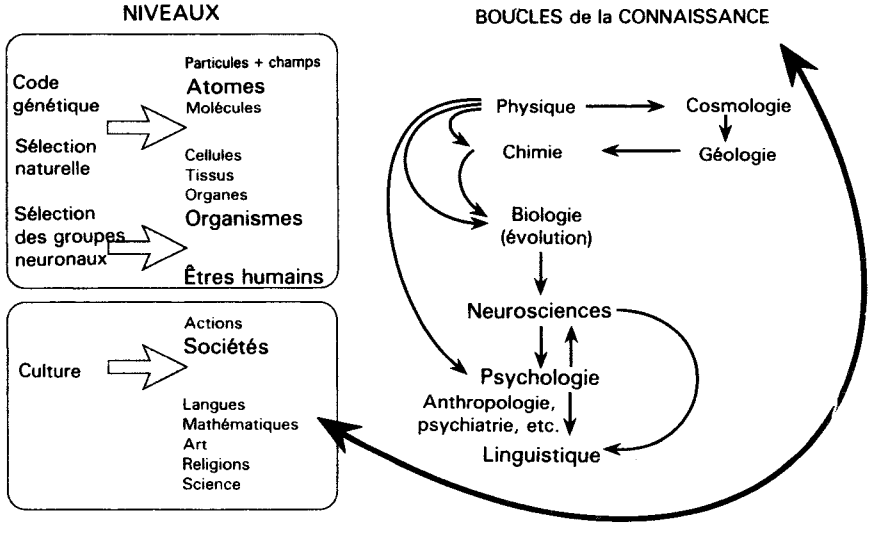


FIGURE 14-1

Les niveaux de l'organisation biologique et les boucles de la connaissance. Voir les échelles respectives sur la figure P-1.

Ces penseurs pourront être déçus en apprenant que les réponses à bien des problèmes fondamentaux concernant l'esprit viendront de l'analyse de la complexité de sa structure, qui est régie par des

principes d'organisation originaux. Mais à y regarder de plus près, cette alternative est riche et pleine de surprises, et elle est plus à même de former un tout avec les splendides théories compatibles de l'évolution et de la physique !

Nous commençons à peine à démêler cette complexité, et le présent effort de synthèse paraîtra sans doute insignifiant une fois achevé. Mais même aux stades les plus précoces, nous avons besoin de disposer d'une vision globale de la question de l'esprit pour pouvoir progresser. Ni une simple comparaison analogique, ni le rationalisme, ni la physique à elle seule ne suffiront. Il faut disposer d'une théorie de la matière elle-même. Celle qui est décrite ici vise à stimuler la construction par d'autres chercheurs de théories alternatives soumises aux mêmes contraintes. Et il ne s'agit pas d'énoncer des hypothèses philosophiques, ou des formulations de haut niveau, mais des théories biologiques qui remettent en cause les faits que j'ai présentés ou l'interprétation que j'en ai donnée.

En attendant ces nouvelles théories, j'espère que le lecteur accueillera avec bienveillance la description sommaire que je m'apprête à faire des différents niveaux et des différentes boucles. Je suppose qu'au point où nous en sommes, il est en tout cas suffisamment familiarisé avec l'étrange vocabulaire que j'emploie pour que je puisse m'en servir ici afin de résumer mon point de vue. Je vais procéder en sens inverse – en partant de la théorie adoptée pour aboutir à ma critique des différentes alternatives. Je vous prie de bien garder présente à l'esprit la notion très générale qui consiste à dire qu'en tant que système sélectif, le cerveau (et tout particulièrement le cortex cérébral) est un corrélateur. Il établit des corrélations entre des entrées temporelles au cours de son propre développement et, au cours de son fonctionnement à l'âge adulte, des corrélations entre les propriétés des signaux et des scènes pour donner naissance à la conscience.

Ma première hypothèse est que la conscience est apparue à l'issue d'une sélection naturelle. L'existence et le fonctionnement de l'esprit dépendent de la conscience. Une notion apparentée est que la conscience est efficace, c'est-à-dire qu'elle fait croître l'adaptation à certains types d'environnements. La conscience naît d'un ensemble particulier de relations existant entre la perception, la formation des concepts et la mémoire. Ces fonctions psychologiques dépendent des mécanismes de catégorisation existant dans le cerveau. De plus, la mémoire est influencée par des systèmes de valeurs établis au

cours de l'évolution et par des systèmes de contrôle homéostatique caractéristiques de chaque espèce.

La conscience primaire est obtenue par réentrée d'une mémoire des valeurs-catégories dans les catégorisations perceptives simultanément en cours dans de nombreuses modalités. Elle joint dans le temps et dans l'espace les stimulus qui surviennent en parallèle (y compris ceux qui ne sont pas nécessairement liés de façon causale) en une scène corrélée. Chaque animal confère aux caractéristiques de cette scène un relief qui dépend des valeurs et de l'apprentissage passés de cet animal. La conscience primaire est confinée au présent remémoré. Elle est nécessaire à l'apparition de la conscience d'ordre supérieur, et elle continue à opérer chez les animaux dotés d'une conscience d'ordre supérieur.

La conscience d'ordre supérieur naît avec l'apparition de compétences sémantiques au cours de l'évolution des espèces, et elle s'épanouit avec l'acquisition du langage et de références symboliques. Les compétences linguistiques requièrent un nouveau type de mémoire pour la production et l'audition des sons co-articulés rendus possibles par l'apparition d'une chambre supralaryngée (voir la figure 12-1). Les aires de la parole qui interviennent dans la catégorisation et dans la mémoire linguistiques interagissent avec des aires conceptuelles déjà existantes du cerveau. Leur fonction consiste à relier la phonétique à la signification au sein d'une communauté linguistique, en faisant appel à des interactions avec les aires conceptuelles du cerveau pour guider l'apprentissage. Lorsque ces mêmes centres conceptuels effectuent des catégorisations sur l'organisation verbale qui se met en place au cours des actes de parole, cela donne naissance à une syntaxe. À mesure que la syntaxe commence à se construire et qu'un lexique suffisamment étendu est acquis à travers l'apprentissage, les centres conceptuels du cerveau commencent à traiter les symboles et leurs références, ainsi que l'imagerie qu'ils évoquent, comme un monde « indépendant » devant subir des catégorisations supplémentaires. Une explosion conceptuelle et une révolution ontologique – nous nous représentons un monde, et pas simplement un environnement – sont donc rendues possibles par l'interaction des centres conceptuels et linguistiques. C'est ainsi que les concepts de moi, de passé et de futur apparaissent.

La conscience d'ordre supérieur dépend de la construction d'un moi à travers des échanges affectifs intersubjectifs. Ces interactions – avec des figures parentales, avec des congénères au cours de la

toilette et avec des partenaires sexuels – sont du même type que celles qui guident les échanges sémiotiques et la construction du langage. À travers les symboles, des échanges à coloration affective déclenchent un *bootstrapping* sémantique. Il en résulte un modèle d'un monde plutôt que d'une niche écologique, ainsi que des modèles du passé, du présent et du futur. Cependant, tandis que la conscience d'ordre supérieur nous libère de la tyrannie du présent remémoré, la conscience primaire continue de coexister et d'interagir avec les mécanismes de la conscience d'ordre supérieur. En fait, la conscience primaire fournit une forte impulsion aux processus d'ordre supérieur. Nous vivons sur plusieurs niveaux à la fois.

Les propriétés jamesiennes de ces processus conscients (voir le chapitre 11) résultent des fonctions du cortex cérébral et de ses appendices. Ces derniers sont les organes de la succession – le cervelet pour le déroulement sans accrocs des mouvements, l'hippocampe pour l'établissement de la mémoire à long terme et les ganglions de la base pour le choix des configurations de comportements moteurs et des plans d'attention. Leur fonctionnement dépend des déplacements et des actes de l'organisme lorsque celui-ci explore son environnement.

Les propriétés résultantes de subjectivité, d'intentionnalité, de continuité et de changement entrent en jeu ensemble, en apparent unisson. Ces propriétés peuvent s'expliquer dans le cadre de la TSGN étendue sans qu'il soit nécessaire de faire appel à d'autres hypothèses hormis celles concernant la sélection qui s'opère au cours du développement, de la sélection qui s'opère à travers l'expérience, et de la réentrée. De nouvelles fonctions, dont la conscience, sont rendues possibles par de nouvelles morphologies, produites par l'évolution, et reliées par de nouvelles connexions à des structures cérébrales déjà existantes.

La science et le langage « objectifs » dépendent tous deux de la métastabilité, ou constance, des objets du monde physique. La théorie de la conscience suppose que la physique et l'évolution, lorsqu'on y adjoint les hypothèses de la TSGN, suffisent pour construire une science de l'esprit. Cependant, il est impossible de construire une théorie scientifique de l'esprit d'un individu donné, de même qu'il est impossible de décrire scientifiquement tous les événements de l'histoire du monde.

Pour rester scientifique, la TSGN étendue doit supposer qu'à la fois le sujet humain et l'observateur scientifique humain qui étudie ce sujet éprouvent des sensations. Cette hypothèse est nécessaire

pour garantir que les échanges scientifiques intersubjectifs aient un sens. D'après la théorie, les sensations sont des catégorisations, effectuées par la conscience d'ordre supérieur, des « scènes » et des « souvenirs » fournis par la conscience primaire. Elles font intervenir des relations de recatégorisation qui, en dernière analyse, sont régies par la façon dont les valeurs sélectionnées par l'évolution interagissent avec la mémoire.

Les créatures qui ne possèdent qu'une conscience primaire ne peuvent ni décrire des sensations ni y réfléchir. Si tant est qu'elles en ressentent (et nous ne pouvons que conjecturer que ce soit le cas), elles ne le font que dans le présent remémoré. En nous fondant sur des comparaisons morphologiques, nous pouvons envisager l'apparition de trois niveaux de propriétés sensorielles au cours de l'évolution des animaux possédant des neurones :

1. Réactions à des stimulus, les réactions d'aversion et les réactions consommatoires étant directement régies par la sélection de valeurs au cours de l'évolution. Le homard, par exemple, est capable d'apprendre et d'avoir une mémoire à long terme, mais ne possède pas de conscience primaire.

2. Stimulus suscitant des réactions chez des animaux dotés d'une conscience primaire. Une vie mentale faite de scènes jamesiennes qui établissent des corrélations entre des valeurs et des catégorisations perceptives, mais pas de moi socialement constitué. Des bases anatomiques pour les sensations et pour leur discrimination selon différentes modalités. Pas de catégorisation des sensations au cours du temps par un *sujet*, mais existence d'une mémoire à long terme (non consciente en tant que telle) fondée sur des sensations éprouvées dans le présent remémoré. Exemple : les chiens.

3. Stimulus significatifs du point de vue de l'aversion et de l'appétence transformés en un monde, et non simplement en une niche écologique, par des animaux ayant une conscience d'ordre supérieur. Des sensations à part entière, susceptibles d'être affinées, remémorées, modifiées et décrites, comme dans le cas des taste-vin. Exemple : les êtres humains. Exemple extrême : la sainteté – avec un renoncement à tous les impératifs biologiques et notamment des réactions inhabituelles à des sensations douloureuses –, fondée sur une conviction profonde.

Pour l'instant, nous ne pouvons que spéculer sur des questions comme celles-ci. Mais nous savons néanmoins que la conscience d'ordre supérieur conduit à la construction d'un domaine de l'imagination, d'un domaine des sentiments, des émotions, des pensées,

de l'imaginaire, du moi et de la volonté. Elle construit des objets artificiels qui sont des objets mentaux. Du point de vue culturel, ces actes mènent à l'étude des relations stables entre les choses (la science), des relations stables entre des objets mentaux stables (les mathématiques), et des relations stables entre des phrases qui s'appliquent à des choses et à des objets mentaux (la logique). L'incomplétude de ce type de domaines – caractère qui, en mathématiques, fut mis en évidence par Kurt Gödel – pourrait s'expliquer par le fait que la construction de formes dans l'esprit fait toujours appel aux *bootstrappings* d'ordre supérieur qui sont nécessaires à la conscience. Nous pensons par configurations synthétiques, et non en termes logiques, et c'est pour cette raison qu'il se peut que notre pensée aille toujours plus loin que des relations syntaxiques, mécaniques.

L'analogie entre esprit et ordinateur est erronée à de nombreux égards. En particulier, le cerveau est bâti selon des principes qui assurent la diversité et la dégénérescence. Contrairement à un ordinateur, il n'a pas de mémoire réplicative ; il est historique et gouverné par des valeurs. Il forme des catégories d'après des critères internes et des contraintes qui agissent à de nombreuses échelles, et non pas par l'intermédiaire d'un programme construit syntaxiquement. Le monde avec lequel le cerveau interagit n'est pas fait de catégories classiques, définies sans équivoque. (Il est cependant exact que certains « objets naturels » semblent correspondre à ces catégories à cause des caractéristiques interactives de notre phénotype et des propriétés physiques de ces objets.)

Le monde, par conséquent, n'est pas comme un de ces fragments de bande magnétique qui sont lus par les ordinateurs. Et la physique, dont l'objectif est d'étudier un tel monde, décrit ses propriétés corrélatives formelles mais ne contient pas de théorie des catégories permettant d'effectuer une seule et unique partition des objets macroscopiques. Comme je le fais remarquer dans la postface, l'objectivisme échoue.

Les mécanismes de catégorisation opèrent à travers des cartographies globales qui font nécessairement intervenir notre corps et notre histoire personnelle. Par conséquent, la perception n'est pas nécessairement véridique (voir, par exemple, le triangle de Kanizsa, dans la figure 4-2). Notre comportement est régi par une mémoire résultant de recatégorisations effectuées sous l'influence des modifications dynamiques d'un certain nombre de valeurs. Les croyances et les concepts ne sont individualisés que par référence à un envi-

ronnement non figé, dont la description ne peut être spécifiée à l'avance. Nos modes de catégorisation et notre utilisation de métaphores dans nos pensées (le fait d'établir une correspondance entre deux choses appartenant à des domaines différents) reflètent ces observations.

Dans la postface, je fournis des arguments détaillés à l'encontre du point de vue des sciences cognitives, selon lequel l'esprit serait fondé sur des représentations algorithmiques, sur des calculs. L'idée qu'il existe des représentations mentales censées être structurées du point de vue syntaxique (en un « langage de la pensée »), puis plaquées sur un modèle sémantique vaguement spécifié, ou sur un modèle objectiviste soumis à de trop nombreuses contraintes, est incompatible avec les données de l'évolution. Les propriétés proposées par ces modèles cognitifs sont incompatibles avec les propriétés des cerveaux, des corps et du monde. Le but de la TSGN étendue est d'expliquer comment s'incarne l'esprit, reliant ainsi la cognition à la biologie. Elle fournit des bases cohérentes permettant d'expliquer comment l'incarnation donne naissance à la signification à l'issue d'interactions référentielles. Tout ce riche domaine d'étude, qui s'intéresse à la correspondance précise existant entre nos concepts et notre corps, n'en est actuellement qu'à ses balbutiements.

Pourquoi ai-je rejeté l'idée que des systèmes axiomatiques ou syntaxiques puissent être à la base de l'esprit, malgré leur apparente élégance ? Après tout, il est vrai que les systèmes axiomatiques *semblent* souvent fournir de bonnes indications sur la façon dont fonctionne l'esprit, surtout lorsqu'ils sont utilisés en même temps que la physique. Mais en fait, ce ne sont que des constructions sociales qui *résultent* de la pensée, et non la base de cette pensée. Leurs racines se trouvent dans la logique mathématique du XIXe siècle. Ces systèmes ont foisonné grâce à David Hilbert, ils ont été modulés et circonscrits par Gödel, et ils sont souvent conçus sous une forme typologique, essentialiste. Ils ne constituent *pas* un bon modèle de l'esprit, car l'esprit doit les *précéder* afin de les créer et de les faire fonctionner. La conscience est essentielle à leur formulation ainsi qu'au platonisme que parfois ils inspirent, mais les faits démontrent que la conscience est apparue à l'issue d'une évolution, et non d'une typologie. Darwin avait raison : c'est la morphologie qui a donné l'esprit. Et sur ce point, Wallace, qui pensait que la sélection naturelle ne pouvait pas rendre compte de l'esprit humain, avait tort.

Quant à Platon, il n'avait *même pas* tort : il était tout simplement à côté de la question.

Il peut être utile d'évoquer ici un point trivial : le fait que l'évolution de la conscience a dépendu de certaines températures. La stabilité de tous les objets physiques que la science et le chercheur scientifique conscient décrivent est la même que celle qui consolide les phénomènes dénominatifs. Il n'y aurait jamais eu de conscience à 10^6 °C. La conscience est apparue à moment donné, en un lieu donné, et à une température beaucoup plus basse – à une température autorisant les réactions chimiques. Dire cela revient à rejeter le panpsychisme en tant que théorie de l'esprit. (Je m'attarderai davantage sur cette question au chapitre 20, qui traite des origines premières de l'esprit.)

La conscience est au cœur du comportement humain, de la société, du langage et de la science. Si vous envisagiez le contraire, vous seriez obligé de postuler l'existence d'une bande magnétique figée contenant le monde, d'un « cerveau-ordinateur », et d'un « programmeur du monde » à vous faire mourir d'ennui. La TSGN, au contraire, avec ses niveaux et ses boucles complexes, semble mieux s'aligner avec les faits de la biologie et paraît préférable parce qu'elle correspond beaucoup mieux à notre propre expérience.

Fort de cet aveu franc quant à mes préférences personnelles, je vais aborder à présent la dernière partie de ce livre. Je l'ai intitulée *Harmonies* afin de mettre en valeur les interactions fructueuses qu'une science de l'esprit doit entretenir avec la philosophie, la médecine et la physique. Ces disciplines sont toutes différentes, mais après tout, les harmonies les plus intéressantes se fondent sur la consonance d'entités différentes plutôt que sur l'identité ou l'unisson. Nous espérons tous parvenir à réconcilier des points de vue contradictoires, à clarifier des pensées, à harmoniser des idées. Je ne fais pas exception à la règle, et je n'ai pas l'intention de perdre une occasion de philosopher, premièrement, sur la philosophie elle-même, deuxièmement sur le moi, sur ses pensées et ses perturbations, troisièmement sur la possibilité de fabriquer des objets conscients et quatrièmement sur les grands thèmes de la science de demain qui permettra un jour de mettre plus clairement en évidence les liens existant entre la physique et la psychologie.

Dans *Modes of Thought*, Whitehead faisait remarquer que la philosophie constitue une tentative pour rendre manifestes les données fondamentales concernant la nature des choses. Dans le même ouvrage, il remarquait aussi que le raisonnement scientifique est

totalement sous l'emprise du présupposé selon lequel les mécanismes mentaux ne font pas vraiment partie de la nature. Il le déplorait et espérait qu'un lien adéquat entre les phénomènes mentaux et physiques pourrait être construit au sein de la science elle-même. C'était en 1933. Aujourd'hui, dans les années 1990, nous allons peut-être enfin réussir à entrevoir comment forger ce lien sans refermer la porte sur la philosophie, qui est, par-dessus tout, un état d'esprit.

QUATRIÈME PARTIE

Harmonies

Dans cette dernière série de chapitres, nous allons nous demander quelles sont les conséquences de notre nouvelle théorie du cerveau en ce qui concerne les êtres humains (et parfois non humains). Nous verrons que ces chapitres plaident en faveur de l'ouverture d'esprit à propos des questions de l'esprit. Nous verrons aussi qu'ils incitent à penser que nos connaissances ne sont pas immuables, que nous sommes profondément inscrits à la fois dans la matière du monde et dans celle de l'esprit, que chacun de nous est un individu unique (et important), que le fait que notre pensée s'inscrive dans une culture est essentiel à notre humanité et à notre capacité de comprendre le sens des choses et que, même en cas de maladie, notre esprit est merveilleusement adaptable. Nous verrons également qu'on peut penser qu'un jour, qui n'est peut-être pas si désespérément lointain, nous serons capables de construire des objets possédant certaines de nos propriétés psychologiques.

Surtout, cette partie du livre incite à penser que la construction d'une théorie adéquate du cerveau promet de servir à fonder des harmonies nouvelles – y compris celles qui devraient nous permettre de trouver notre place dans l'univers. Ainsi, au cours du dernier de ces chapitres, j'essayerai de répondre à la question suivante : s'il fallait choisir deux grandes idées scientifiques, deux grands concepts susceptibles à eux deux de nous aider à comprendre ce que nous sommes et où nous nous plaçons dans l'ordre des choses, lesquels choisirait-on ?

Chapitre 15

La philosophie et ses affirmations : un cimetière aux « ismes »

> *Deux philosophes arrivent toujours à se dire tout ce qu'ils savent en deux heures.*
>
> *Je ne vois pas pourquoi un homme devrait désespérer parce qu'il ne voit pas de barbe à son cosmos. S'il pense que c'est lui qui se trouve dans le cosmos, et non pas le cosmos qui se trouve en lui, il sait que la conscience, la finalité, la signification et les idéaux font partie des potentialités de ce cosmos* [...] *et c'est l'affaire de la philosophie de montrer que nous ne sommes pas fous de faire ce que nous voulons faire.*
>
> Oliver Wendell Holmes, Jr.

La philosophie s'intéresse depuis toujours à l'esprit et à ses mécanismes. En passant en revue les concepts de la philosophie, Arthur Danto a défini presque toutes les positions philosophiques en fonction de ce qu'il appelle un épisode cognitif de base. Cette notion ramène à Descartes et s'exprime sous la forme d'une relation entre trois composants : un sujet, une représentation et le monde. La relation existant entre le monde et le sujet est la relation de causalité, entre le monde et la représentation celle de vérité, entre le sujet et la représentation celle que le sujet entretient avec lui-même. Danto appelle les humains des êtres représentants ou « représentationnels », et je pense qu'il tombe dans le piège dont nous avons déjà vu les dangers lorsqu'il propose l'idée selon laquelle le corps serait « propositionnellement structuré », signifiant par là que la tâche des neurosciences consiste à « montrer comment le tissu nerveux représente ». Mais si on ne s'en tient pas strictement à ce point de vue, qui se rapproche dangereusement des notions de codage et d'instruction, la triade de Danto s'avère utile, car elle peut nous aider à voir comment, en tant qu'individus, *nous* nous (et non pas nos cerveaux) « représentons » le monde.

Pour un scientifique, la philosophie est parfois déconcertante. En effet, la science est censée fournir une description des lois de ce monde et des façons éventuelles de les appliquer, tandis que la philosophie n'a pas de thème qui lui soit propre. Au contraire, son

but est d'analyser la clarté et la cohérence des autres domaines de la connaissance. De plus, contrairement à la science, on peut dire qu'elle manque de modestie. Il n'y a pas de philosophie partielle ; chaque philosophe formule une théorie complète. Comme un enfant qui tout à coup se met à comprendre le langage, le philosophe ne se doit pas simplement de décrire un environnement donné ; il se doit de construire entièrement le monde. À chaque nouvelle tentative de construction philosophique, on trouve en toile de fond une certaine vision du monde, qui est personnelle. Le fait que toutes ces visions portent des noms qui s'achèvent en « isme » donne un intéressant ensemble (tableau 15-1) – correspondant à l'ensemble de façons dont la triade de Danto a été disséquée, en fonction de l'importance accordée à ses relations internes. Je ne m'attarderai pas ici sur toutes ces dissections ; la description sommaire figurant au chapitre 4 devrait vous suffire à vous faire une idée sur quelques-unes d'entre elles. Cependant, certains lecteurs pourront trouver amusant de comparer leurs ramifications respectives dans un dictionnaire ou une encyclopédie de philosophie. Ils s'apercevront alors que, malheureusement, chacun de ces « isms » est susceptible d'impliquer le rejet du précédent, étant donné que chaque philosophe construit sa vision personnelle et unique du monde. En fait, la philosophie est un véritable cimetière aux « ismes ».

TABLEAU 15-1

Quelques « ismes » philosophiques *

Empirisme	Monisme	Réalisme
Rationalisme	Dualisme	Idéalisme
Phénoménalisme	Pluralisme	
Réductionnisme	Epiphénoménalisme	Essentialisme
Objectivisme	Matérialisme	Behaviorisme
Opérationnalisme	Panpsychisme	(Behaviorisme philosophique)
Instrumentalisme	Déterminisme	Représentationnalisme
Positivisme logique	Compatibilisme	Fonctionnalisme
Fondationnalisme	Incompatibilisme	Interactionnisme
Pragmatisme	Occasionnalisme	Internalisme
Evolutionnisme		Externalisme
Sélectionnisme		Existentialisme

* *Cette liste pourrait prendre des proportions indécentes si l'on y incluait également les idéologies morales, esthétiques, cliniques, religieuses et politiques. Les ismes les plus décidément non scientifiques sont notamment le géocentrisme, le vitalisme et le mécanisme. Bien entendu, les doctrines ne sont pas toutes des ismes (la réciproque étant probablement vraie elle aussi), mais elles sont susceptibles de le devenir. Voilà le danger.*

Dans ce cas, pourquoi s'en préoccuper ? Parce que la philosophie tente de réfléchir à *tous* les aspects de notre existence individuelle et collective ; parce que son histoire est intimement liée à celle de la psychologie ; et aussi parce que la mise au point d'une nouvelle vision scientifique de l'esprit, fondée sur la biologie, pourrait aider la philosophie à prendre un nouveau départ dans la vie.

Dans leur livre intitulé *The Anthropic Cosmological Principle*, John Barrow et Frank Tipler attribuent à un président d'université anonyme la citation suivante :

> Pourquoi faut-il que vous, les physiciens, vous ayez toujours besoin d'équipements si coûteux ? Les mathématiques, voilà un département qui n'a pas besoin d'acheter autre chose que du papier, des crayons et des corbeilles à papiers. Et le département de philosophie est encore mieux : il n'a même pas besoin de corbeilles à papiers.

Songez aussi que, selon Einstein, le principal outil du physicien théorique n'est autre que la corbeille à papiers !

La philosophie du XXe siècle se caractérise par l'abandon des grands objectifs passés de synthèse, objectifs déjà évoqués au chapitre 4. Depuis Wittgenstein, une bonne part de la philosophie a été consacrée à mettre de l'ordre dans la logique et le langage. Depuis Edmund Husserl, une autre s'est intéressée à des considérations délibérément non scientifiques sur la conscience et l'existence – à ce qu'on appelle la phénoménologie. Il vaut donc certainement la peine de se demander si une théorie de l'esprit fondée sur la biologie pourrait ou non donner du tonus à ces domaines de la pensée, et même, peut-être, redonner une nouvelle tournure à la philosophie.

Entrons dans le jeu, et voyons combien d'ismes tombent si nous adoptons un point de vue scientifique sur l'esprit. Bien sûr, il nous faudra également dire quelque chose sur les limites de la science et de la connaissance elle-même. Mais énonçons d'abord les hypothèses de base de toute vision scientifique :

1. Il existe un monde réel – un monde décrit par les lois de la physique, qui s'appliquent partout. (C'est l'hypothèse physique.)

2. Nous sommes partie intégrante de ce monde, nous suivons ses lois et nous sommes le produit d'une évolution à partir d'une origine ancienne. C'est l'apparition, au cours de l'évolution, de nouvelles morphologies qui a donné naissance à l'esprit. (C'est l'hypothèse évolutionniste.)

3. Il est possible de replacer l'esprit dans la nature. Il est possible de construire une science de l'esprit sur des bases biologiques. Et la façon d'éviter les cercles vicieux et les impasses consiste à construire une théorie du cerveau fondée sur des principes sélectionnistes. (C'est là l'argument central de ce livre.)

Si nous acceptons ces hypothèses et les raisonnements que nous avons tenus jusqu'ici dans ce livre, nous pouvons immédiatement enterrer quelques ismes de plus – le dualisme, le panpsychisme, l'épiphénoménalisme, l'idéalisme, le représentationnalisme, l'empirisme et l'essentialisme. Ils sont incompatibles à la fois avec les hypothèses ci-dessus, avec les données issues de la psychologie et des neurosciences, et avec la biologie elle-même. Je ne m'étendrai pas sur les raisons qui permettent de l'affirmer. J'énoncerai simplement ces quelques phrases présumées fatales dans l'ordre des ismes que je viens de citer : il n'y a pas de *res cogitans* ; les particules ne sont pas conscientes ; la conscience est efficace du point de vue évolutif ; le monde existe et persiste indépendamment de l'esprit, et il existait déjà avant lui ; le cerveau est un système sélectif et non pas une machine de Turing ; les données sensorielles ne constituent pas la base de l'esprit ; le « monde » n'est pas composé de catégories classiques ; la typologie est détruite par la biologie. Disons également qu'au cours des trois cents dernières années, la science a déjà détruit d'autres idées plus sectaires, comme le géocentrisme, le vitalisme et le mécanisme simple.

Trêve de destruction, du moins pour l'instant. Que pouvons-nous dire de constructif sur la science et sur la possibilité de bâtir une théorie de la connaissance fondée sur la biologie – une épistémologie biologiquement fondée ?

Afin de construire une description raisonnable, nous devons reconnaître que la physique moderne des particules et la théorie des champs ont aboli l'idée d'un monde déterministe, fonctionnant comme un mécanisme d'horlogerie. Cela ne signifie pas cependant que les mécanismes ne puissent être décrits ou qu'ils soient inutiles (ils sont même très utiles en physique et en biologie), mais tout simplement qu'il n'est pas raisonnable de considérer l'univers en ces termes *à toutes les échelles* (voir figure la P-1). Nous devons également reconnaître que la théorie darwinienne de l'évolution a assené un coup mortel à l'essentialisme et au platonisme. Enfin, nous devons tenir compte du fait que, au cours de l'évolution, les systèmes de sélection naturelle (qui sont des systèmes

historiques) ont donné lieu à des systèmes sélectifs somatiques capables de gérer les nouveautés survenant au cours de la vie d'un individu.

Ce dernier élément est moins bien établi que les autres – l'idée qu'il existe des sciences de la reconnaissance, au centre desquelles se trouvent notamment les neurosciences, n'a pas encore été généralement admise en ce qui concerne l'étude du cerveau. Mais si nous admettons que les principales prémisses du darwinisme neuronal sont correctes (et les données en leur faveur ne cessent de s'accumuler), nous pouvons parvenir à plusieurs conclusions intéressantes. La première est que nous n'avons pas besoin de nous aventurer au-delà de la biologie elle-même à la recherche d'explications exotiques de l'esprit. Rappelez-vous que les hypothèses de la TSGN – sélection et variabilité au cours du développement, sélection synaptique, et amplification différentielle au sein des systèmes réentrants – sont *les seuls* principes proposés par cette théorie. Autrement dit, il n'est pas nécessaire d'adjoindre aucun principe nouveau pour rendre compte de la conscience – il ne faut que de nouvelles morphologies, issues de l'évolution. La deuxième conclusion est que ces concepts, s'ils sont exacts, excluent toute possibilité de décrire les mécanismes cérébraux de manière *globale* en termes de machines de Turing, d'ordinateurs. Et la troisième est que, bien que le dualisme des substances (la version cartésienne du dualisme) et le dualisme des propriétés (l'idée que la psychologie ne peut être décrite de manière satisfaisante qu'en termes psychologiques) soient exclus, nous devons admettre qu'il existe une distinction entre les systèmes matériels sélectifs et non sélectifs. D'après cette distinction, les systèmes vivants et mentaux sont sélectifs. Autrement dit, il existe une réelle distinction entre la biologie (ou la psychologie) et la physique. Ainsi, tout en acceptant que les lois de la physique s'appliquent à tous les systèmes – qu'ils soient ou non doués d'intentionnalité –, ce point de vue conduit à rejeter l'idée selon laquelle une théorie physique suffisamment sophistiquée (telle la gravitation quantique ou d'autres concepts spécifiques de la physique fondamentale) serait nécessaire pour parvenir à expliquer l'esprit.

Où en sommes-nous à présent avec les ismes ?

En acceptant l'idée d'une épistémologie fondée sur la biologie, nous sommes, en un certain sens, des réalistes, et aussi des matérialistes sophistiqués. En effet, les données issues du développement

et de l'évolution nous amènent à rejeter la téléologie (la doctrine des causes finales, des buts ultimes). Mais, en même temps, nous admettons que l'évolution *peut* sélectionner des animaux de telle sorte qu'ils aient des objectifs généraux, des buts et des valeurs, bref qu'ils incarnent ce qu'on appelle parfois des systèmes téléonomiques. Comme le disait Mayr, la téléonomie est une prédiction du passé. De fait, l'expérience passée de sélection naturelle règle l'ensemble des points fixes des systèmes de valeurs (par exemple, ceux associés à la faim, à la soif et aux réactions sexuelles) de façon à améliorer les chances de survie. Dans notre cas, le cerveau de l'être humain conscient, agissant en tant que système sélectif somatique, utilise les contraintes imposées par les valeurs pour projeter le futur en fonction de catégories et d'objectifs. Ainsi, en supposant que le cerveau est un système sélectif somatique, nous excluons l'idée de l'existence d'un homuncule dans notre tête. Celui-ci n'est pas plus nécessaire aux sciences de la reconnaissance somatique qu'une création particulière ou l'idée de finalité ne le sont à l'évolution. S'il est *res cogitans*, il est exorcisé.

L'esprit, qui est né dans des systèmes matériels mais qui peut pourtant poursuivre des objectifs et des finalités, n'en est pas moins le produit des processus historiques et des contraintes fondées sur des valeurs associées à l'évolution. Quelles limites cela impose-t-il à notre connaissance et à notre liberté ?

Le fait que l'origine de la conscience soit d'ordre matériel ne restreint pas le commerce intellectuel ; la philosophie elle-même en témoigne. Mais cela nous impose certainement des limites, en dépit de notre capacité de construire des dispositifs physiques capables de prolonger nos sens et de faire augmenter notre puissance de calcul. D'après la définition que nous avons donnée de la signification dans ce livre, nous sommes forcés d'admettre que notre position est celle de réalistes *pondérés*. Notre description du monde est en effet pondérée par la façon dont se forment nos concepts. Et, en dépit du fait qu'il pourrait exister une liberté infinie *au sein* d'une grammaire, notre langage et nos concepts sémantiques vont bien au-delà des règles grammaticales. J'ai déjà décrit la façon dont je pense que la signification naît de l'incarnation, à travers la sélection de groupes neuronaux et la réentrée. Ainsi, malgré les remarquables extensions sémantiques permises par nos calculs et nos expériences, nous sommes obligés d'admettre qu'il se pourrait bien que nous soyons limités dans notre pensée par la façon dont

nous sommes constitués, c'est-à-dire par le fait que nous sommes les produits d'une évolution morphologique.

Nous pouvons adjoindre trois autres éléments importants à cette vision du réalisme pondéré et de l'épistémologie biologiquement fondée. Ces trois éléments sont : (1) l'existence d'un ensemble extraordinairement dense d'événements dans le monde réel ; ainsi, malgré notre aptitude à en classer de grandes quantités par catégories, nous pouvons difficilement en donner une description exhaustive ; (2) beaucoup d'événements sont irréversibles ; et (3) chez chaque individu, les sensations et les perceptions suivent des cours différents, irréversibles et relevant de l'idiosyncrasie.

D'après l'épistémologie biologiquement fondée et le réalisme pondéré, la connaissance *doit* rester fragmentaire et sujette à correction. La certitude cartésienne n'existe pas. Mais que dire alors, pourrez-vous objecter, des certitudes mathématiques, des relations analytiques et des tautologies ? Il ne m'appartient pas ici de considérer ces questions en profondeur. Cependant, il est utile de dire clairement que les systèmes de ce type sont artificiels, et qu'ils ont été créés par l'esprit à travers des interactions sociales et des manipulations de symboles par des individus. De plus, Gödel a démontré que même le plus fondamental de ces systèmes, l'arithmétique, est incomplet. Pour moi, l'étude des mathématiques peut être définie – comme l'ont si bien dit Philip Davis et Reuben Hersh dans *The Mathematical Experience* – en tant qu'étude d'objets mentaux stables ou invariants. Bien que notre subtil matérialisme soit loin d'affirmer que ces relations sont dépourvues de sens, il leur interdit d'avoir une existence platonique indépendante.

Nous avons déjà évoqué certaines autres limites. Par exemple, nous avons vu qu'étant donné les limitations pesant sur notre savoir ainsi que notre état d'êtres « fermés de l'intérieur », les détails du parcours personnel et irréversible des sensations chez un individu donné ne sont pas accessibles à un autre individu. (En fait, ce parcours peut même devenir inaccessible à son propre détenteur.) De plus, nous devons admettre que la mort signifie la perte irrévocable d'un individu et de son être. La mort n'est pas une expérience, car il n'y a alors plus rien à signaler. Les esprits désincarnés n'existent pas. Toutes ces limites à notre connaissance et à la science sont évidentes. Mais cela ne nous empêche pas de tenter de définir clairement notre place dans la vision du monde fournie par la science.

À la nécessité de pondérer notre réalisme s'ajoute celle d'aborder

certaines questions historiques et culturelles, ainsi que d'autres liées aux valeurs et à la finalité. Cela peut paraître étrange dans une réflexion sur la science, car la science est censée être indépendante de toute valeur. Mais en fait, la seule science présentée comme indépendante des valeurs est celle qui se fonde sur la perspective galiléenne, une science physique qui, de façon tout à fait délibérée et légitime, a placé l'esprit hors de la nature. Une épistémologie fondée sur la biologie ne peut pas se permettre un tel luxe.

Il est bon de s'attarder quelque peu sur cette question, car elle en dit long sur la place qu'occupe la science dans nos vies. L'expérience humaine consciente a donné lieu à la culture, et la culture à l'histoire. Mais l'histoire n'est pas une simple chronique ; c'est une interprétation qui prend en compte des causes et des valeurs que l'on croit être exactes. La science est apparue au sein de l'histoire, et elle tente de décrire, avec beaucoup plus de certitude, les frontières du monde – ses contraintes et ses lois physiques. Mais ces lois ne peuvent remplacer l'histoire ni le cours particulier de la vie de chaque individu. Un ensemble de lois ne pourra jamais se substituer à l'expérience, et il ne sera certainement jamais équivalent à un ensemble d'événements. Les lois ne décrivent pas – elles ne peuvent pas décrire – l'expérience de façon exhaustive, ni remplacer l'histoire ou les événements qui surviennent effectivement au cours de la vie des individus. Les événements sont plus denses que toute description scientifique. Ils sont également indéterminés au niveau microscopique et d'ailleurs, selon notre théorie, ils sont même indéterminés – dans une certaine mesure – au niveau macroscopique.

Certains pourront penser que je suis en train d'essayer de limiter *a priori* les capacités des descriptions scientifiques. Mais il n'en est rien. Je ne fais qu'indiquer que, même si la science parvient à redonner à l'esprit une place dans la nature, elle ne sera pas capable, d'après ce que nous avons dit, de décrire correctement l'expérience individuelle ou historique. Mais il n'en demeure pas moins qu'elle fournit une description satisfaisante (en fait, la meilleure) des *contraintes* qui pèsent sur l'expérience.

Parmi les contraintes dues au fait d'adopter le point de vue sélectionniste, l'une est issue de la TSGN étendue : aucun système fondé sur la sélection ne fonctionne indépendamment d'un certain ensemble de valeurs. Les valeurs sont des contraintes requises par les mécanismes adaptatifs des espèces. Au sein de notre espèce, par exemple, le fait que les fonctions physiologiques, la faim et le sexe

soient toujours semblables suggère l'existence d'un ensemble de propriétés communes à tous les individus. Le cerveau est structuré de façon à jouer un rôle clé dans la régulation des systèmes de valeurs issus de l'évolution qui sont sous-jacents à ces propriétés. Et ces systèmes sont sans doute également à la base des constructions d'ordre supérieur qui constituent les objectifs et les finalités individuels. Nos catégorisations se fondent sur des valeurs.

À la suite de l'apparition de la conscience d'ordre supérieur, il est devenu possible de modifier les valeurs au niveau biologique – bien que seulement jusqu'à un certain point. Bien sûr, comme je l'ai déjà évoqué, nous devons admettre qu'il est possible que certains organismes – ce que nous appelons des martyrs et des saints – soient capables de renier presque totalement les valeurs biologiques. De fait, seules les créatures dotées d'une conscience d'ordre supérieur peuvent ainsi transcender les diktats de la biologie. Mais si l'on convient de ne pas tenir compte des saints, il est clair que l'insertion d'objectifs, de finalités et de valeurs éthiques dans les systèmes sociaux, aussi éloignés soient-ils des systèmes de valeurs biologiques de base, provient presque certainement du fait que les systèmes sélectifs du cerveau ont besoin d'être guidés par des valeurs. D'ailleurs, quelle que soit la culture, les décisions concernant les valeurs sociales doivent toujours précéder celles qui privilégient l'intérêt de la science, et ce malgré l'importance de la connaissance scientifique.

La science est devenue éminemment concrète, et elle se doit de l'être, puisqu'elle est au service des vérités susceptibles d'être vérifiées. La société moderne et son économie dépendent de plus en plus de la technologie scientifique, et les convictions scientifiques sont assimilées par un nombre de plus en plus considérable de gens. Hormis la curiosité, cependant, c'est souvent la cupidité qui stimule à présent la quête de connaissances. Mais quelles que soient les motivations de cette recherche, on peut dire que la science a ceci de bon qu'elle contribue à transformer utilement nos conditions matérielles de vie (pourvu que nos valeurs restent claires). Cependant, l'équilibre entre le bien-être privé et le bien-être public est toujours instable, comme nous l'a appris tragiquement l'histoire de ce siècle industriel et atomique. Le pouvoir ne rend pas perspicace, et le fait de passer d'une science fondée uniquement sur la physique à une science fondée à la fois sur la physique et la biologie va vraisemblablement nous éclairer de façon beaucoup plus profonde et nous aider à changer la façon dont les décisions sociales impor-

tantes sont prises. En effet, à mesure que notre compréhension du cerveau s'améliorera, l'épistémologie fondée sur la biologie aura sans doute d'importantes choses à dire sur de telles décisions.

Le fait de connaître notre place dans le monde réel grâce à une théorie de l'esprit fondée sur la biologie nous révélera également nos limites et mettra un frein à nos ambitions philosophiques. Mais dans certaines directions, ces limites ne nous imposent pour ainsi dire aucune contrainte. En particulier, l'imagination de l'être humain conscient vivant au sein d'une culture est potentiellement illimitée. En dépit de nos limitations (chacun de nous est enfermé dans sa propre expérience consciente), malgré le fait que nous soyons mortels et que notre réalisme soit pondéré, le futur reste ouvert ; il n'est pas prédéterminé. Nous ne pouvons pas chercher refuge dans le fondationnalisme, dans une philosophie première, et nous ne pouvons pas non plus connaître avec certitude tout ce que nous parvenons à apprécier ou à structurer. Nous sommes pour la plupart incapables de renier les valeurs biologiques que l'évolution a sélectionnées pour nous, et nous n'avons d'ailleurs même pas intérêt à le faire, puisque ces valeurs fournissent une base commune à nos décisions morales. Mais l'histoire des découvertes scientifiques et les exploits de l'imagination humaine nous réservent toujours des surprises et, avec l'ascension des neurosciences, fournissent des bases de plus en plus solides aux efforts que nous déployons pour parvenir à nous intégrer dans notre propre description du monde.

En définitive, nous devons donc conclure que nous n'avons pas réussi à tuer tous les ismes. Au contraire, nous en avons privilégié un certain nombre : le réalisme pondéré, le matérialisme sophistiqué, le sélectionnisme et le darwinisme. Le fait de tenir compte de leur importance et de les relier à ce qu'apportent à la fois la biologie et la physique devrait permettre d'enrichir la philosophie et de garantir son harmonisation avec la science. Après tout, penser ne veut pas dire avoir une théorie de l'esprit, et il faudra encore beaucoup réfléchir aux systèmes sélectifs. Par sa nature même, le point de vue que j'ai adopté concernant la nécessité d'une épistémologie fondée sur la biologie suggère que la phénoménologie non scientifique et les exercices grammaticaux, quelle que soit leur valeur, imposent un ensemble trop étroit de limites à l'aventure philosophique. La philosophie a besoin d'un nouveau souffle.

Je pense que les neurosciences joueront un rôle central dans cette évolution. Mais le fait de réduire les choses à des champs quantiques, à des particules étranges, ou à des choses de ce genre n'aura pas

d'emprise sur elle. La tâche consistera, au contraire, à voir comment les systèmes sélectifs du cerveau, fondés sur des valeurs, donnent naissance à la signification et à l'individualité, et aussi comment le moi interprète les frontières du monde. Et elle aura des retombées aussi bien en physique qu'au niveau de la vision sociale de la valeur des individus. Examinons à présent quelques-unes de ses conséquences.

Chapitre 16

La mémoire et l'âme : non au réductionnisme idiot

La science ne peut pas résoudre l'ultime mystère de la nature. Et cela parce qu'en dernière analyse, nous faisons nous-mêmes partie du mystère que nous essayons de résoudre.

Max Planck

Si je devais vivre ma vie à nouveau, je la vivrais au-dessus d'une épicerie fine [1].

Woody Allen

L'explosion créative qui a marqué le Siècle des Lumières – entre le dernier quart du XVII[e] siècle et la dernière décennie du XVIII[e] – a modifié le cours de l'histoire des idées. De nombreux points de vue prédominèrent à cette époque, mais les principaux furent de loin la raison, la science, la liberté et l'individualité humaines. La science sous-jacente était la physique – le système de Newton – et la philosophie sociale, dans une large mesure, celle de Locke. Pourtant, durant le Siècle des Lumières, les idées de causalité et de déterminisme, alliées à une vision mécaniste de la science, minèrent d'emblée tout espoir de voir apparaître une théorie des actions humaines fondée sur la liberté. En effet, l'idée que nos actes sont déterminés par des forces naturelles – par des mécanismes – est difficilement compatible avec l'idée qu'un individu puisse effectuer librement des choix moraux. De plus, alors que les conceptions de l'époque accordaient une grande importance au rôle de la raison et de la culture dans de tels choix, on n'imaginait pas à quel point l'esprit de tous les êtres humains (y compris celui des plus « raisonnables » – c'est-à-dire des plus « cultivés » d'entre eux) est profondément influencé par des forces inconscientes et des émotions.

Quelles qu'aient été les formes prises par ce point de vue dominant suivant les périodes et les lieux, il reste que la philosophie

1. Dans l'original, *If I had to live over again, I'd live over a delicatessen.* Il s'agit là d'un jeu de mots fondé sur le fait que « over again » signifie « à nouveau » et que, ici, « over » signifie au-dessus (N.d.T.).

des Lumières était une philosophie laïque et qu'elle fut à l'origine d'un grand nombre d'idées sous-jacentes à la démocratie moderne. Mais, en dépit de la valeur de cet héritage, le Siècle des Lumières n'est plus. Le premier coup sérieux porté à ses conceptions furent les attaques de Hume à la fois contre le rationalisme et contre l'idée que le progrès humain était lié aux sciences naturelles. Le principal défaut de la philosophie des Lumières était son incapacité de produire une description scientifique adéquate de l'individu humain pour accompagner sa description d'un univers mécanique. Et, du point de vue social, son échec réside dans le fait qu'elle a été incapable de dépasser le concept d'une société composée d'individus égoïstes et heureux en affaires, mais d'un « humanisme » très superficiel. Les penseurs de l'époque ont bien sûr tenté de nous fournir une vision plus vaste, plus stimulante de nous-mêmes. Mais la science dominante étant la physique mécaniste, il manquait un corps de données ou d'idées leur permettant de relier le monde, l'esprit et la société par un raisonnement scientifique, comme ils aspiraient à le faire. Toutefois, quelles qu'aient été les failles et les incohérences de la philosophie des Lumières, celle-ci nous a légué de grands espoirs concernant la place de l'individu dans la société.

Pouvons-nous espérer mieux faire grâce à une vision scientifique de l'esprit ? Au cours de ce chapitre, je me propose de montrer que le type de réductionnisme qui a voué à l'échec l'entreprise des penseurs du Siècle des Lumières peut être réfuté grâce à des résultats issus à la fois des neurosciences et de la physique modernes. J'ai déjà argué qu'on ne peut rendre compte d'un individu seulement en termes moléculaires, physiologiques ou encore propres à la théorie des champs. Il est tout simplement absurde de vouloir réduire une théorie du comportement individuel à une théorie d'interactions moléculaires. Cela devient évident dès que l'on songe au nombre de niveaux différents d'interactions physiques, biologiques et sociales qui doivent être mis en place avant qu'apparaisse la conscience d'ordre supérieur. Le cerveau comporte 10^{11} cellules et au moins 10^{15} connexions. Chaque cellule possède une biochimie régulatrice extrêmement complexe, régie par un ensemble particulier de gènes. Les cellules se rassemblent au cours de la morphogenèse et échangent des signaux en fonction de leur position dans l'embryon, construisant ainsi un corps et un cerveau équipés d'innombrables boucles de contrôle qui obéissent toutes aux mécanismes homéostatiques dont dépend la survie. La sélection qui s'opère sur les répertoires neuronaux induit des modifications dans des myriades de synapses à

mesure que les cellules meurent ou se différencient. Du fait de la survie et des déplacements de l'animal dans le monde, des catégorisations perceptives et conceptuelles ont constamment lieu dans les cartographies globales. La mémoire interagit dynamiquement avec la catégorisation perceptive par l'intermédiaire de connexions réentrantes. L'apprentissage, qui fait intervenir les liens existant entre la catégorisation et les valeurs (sous sa forme la plus subtile, au sein d'une communauté linguistique), associe les capacités symboliques et sémantiques aux centres conceptuels qui contiennent déjà des structures incarnées destinées à l'élaboration des significations.

Il est presque impossible d'établir des comparaisons entre individus concernant le nombre de combinaisons moléculaires significatives qui interviennent dans une telle séquence d'événements, même lorsqu'il s'agit de jumeaux identiques. De toute façon, cela ne servirait à rien. Les cartographies mettent en correspondance de nombreuses cartes avec de nombreuses autres cartes, et les processus sont individuels et irréversibles. Je me demande ce que les humanistes du Siècle des Lumières auraient pensé de tout cela. En tout cas, ces considérations auraient peut-être plu à Diderot, qui, comme nous l'avons vu au chapitre 3, s'est interrogé sur le système nerveux de son ami dans *Le Rêve de d'Alembert*. Les vues de Diderot sur la conscience humaine ont ouvert la voie à l'idée que le fait d'être humain allait au-delà d'une simple question de physique.

J'ai affirmé qu'il ne pourra exister de science complète, et certainement pas de science des êtres humains, tant que la conscience n'aura pas été expliquée en termes biologiques. Étant donné notre conception de la conscience d'ordre supérieur, cela signifie également qu'il nous faudra décrire les bases sur lesquelles se constitue la personne, l'individualité. Par individualité, je n'entends pas simplement l'individualité qui émerge à travers la génétique ou l'immunité, mais l'individualité personnelle issue des interactions qui ont lieu au cours du développement et de la vie en société.

Ce concept d'individualité est crucial du point de vue philosophique, et il est vrai que certains des problèmes qui lui sont associés pourraient être accentués par le point de vue sélectionniste que j'ai adopté par rapport à l'esprit. Cependant, n'oubliez surtout pas qu'en vertu de notre hypothèse sur les sensations, il est impossible de construire une théorie scientifique du moi individuel. Quoi qu'il en soit, je pense que nous pouvons progresser vers une vision plus complète du concept de liberté individuelle, qui constitue l'une des

notions essentielles à toute théorie philosophique des valeurs humaines.

Les questions que je souhaite aborder concernent la relation existant entre la conscience et le temps, les aspects individuels et historiques de la mémoire, et le fait de savoir si notre conception du sujet conscient et pensant modifie ou non notre notion de causalité. Je voudrais aussi parler brièvement du lien qui existe entre les émotions et nos idées concernant l'incarnation de la signification. En dernière analyse, toutes ces questions ont trait au problème du libre arbitre et donc à l'existence d'une morale en dépit de notre condition de mortels.

D'après la TSGN étendue, la mémoire est l'élément clé de la conscience, qui est intimement liée à la continuité et à différentes échelles de temps. La catégorisation perceptive comporte un élément temporel très net, et l'établissement d'une mémoire fondée sur des concepts fait intervenir un élément temporel encore plus étendu. Les mouvements physiques de l'animal constituent la force motrice de ses catégorisations perceptives, et la création de sa mémoire à long terme dépend des transactions temporelles qui ont lieu dans son hippocampe. Comme nous l'avons déjà vu, on peut rendre compte des propriétés jamesiennes de la conscience à partir des mécanismes mis en œuvre dans ce type d'élément. Mais chez les êtres humains la conscience primaire et la conscience d'ordre supérieur coexistent, et chacune d'entre elles entretient des relations différentes avec le temps. Dans la conscience d'ordre supérieur, la notion du temps passé est *conceptuelle*, et elle est liée à des agencements antérieurs de catégories par rapport à un présent immédiat, qui, lui, est régi par la conscience primaire. Contrairement à la conscience primaire, la conscience d'ordre supérieur ne se fonde pas sur l'expérience en cours, mais sur la capacité de modéliser le passé et le futur. Quelle que soit l'échelle considérée, la notion de temps est d'abord et surtout un événement conscient.

Les idées de conscience et de temps « vécu » sont donc intimement liées. Il est d'ailleurs très révélateur de comparer à ce propos la définition donnée par William James et les réflexions de saint Augustin. En effet, le premier définissait la conscience comme quelque chose dont « nous connaissons [le sens] tant que personne ne nous demande de l'expliciter » ; le second écrit, dans ses *Confessions* : « Qu'est-ce donc que le temps ? Si personne ne me le demande, je le sais ; mais si on me le demande et que je veuille l'expliquer, je ne le sais plus. » La notion de continuité du temps personnel,

historique et institutionnel était au centre de la pensée de saint Augustin.

Le temps fait intervenir des successions. L. E. J. Brouwer, l'un des partisans de l'intuitionnisme en mathématiques, a émis une hypothèse fascinante à propos du lien qui existerait entre le temps et le concept de nombre. Selon lui, tous les éléments mathématiques (et en particulier la suite des entiers naturels) proviennent de ce qu'il appelle la « deux-icité ». La deux-icité est le contraste qui existe entre l'expérience consciente en cours (dont la conscience primaire est l'un des principaux éléments) et la conscience directe des expériences passées (qui requiert l'intervention de la conscience d'ordre supérieur). L'idée est fascinante parce qu'elle suggère que le concept de nombre que chacun de nous possède pourrait ne pas résulter simplement de notre perception de différents ensembles de choses dans le monde extérieur : en fait, il pourrait aussi venir de l'intérieur – de notre intuition de la deux-icité, ou de celle de la deux-icité plus celle de la continuité. En procédant de manière récursive, on parviendrait ainsi à la notion d'entier naturel.

Quelle que soit l'origine de ce type d'abstractions, le sens personnel du sacré, le sens du mystère, le sens de l'existence d'un ordre et celui de la continuité sont tous liés à la continuité temporelle telle que nous la ressentons. Et nous la ressentons en tant qu'individus, c'est-à-dire chacun d'une manière quelque peu différente.

En fait, que ce soit dans la conscience primaire ou dans la conscience d'ordre supérieur, le flux de catégorisations est individuel et irréversible. Il constitue une histoire. La mémoire grandit dans une direction ; et, en présence de moyens verbaux, la notion de durée devient une forme supplémentaire de catégorisation. Cette conception du temps se distingue de la notion relativiste de temps d'horloge utilisé par les physiciens – qui, du point de vue microscopique, est réversible. Mais, outre la variabilité et l'irréversibilité des événements physiques *macroscopiques* reconnus par les physiciens, l'une des raisons profondes de l'irréversibilité du temps en tant qu'expérience individuelle réside dans la nature des systèmes sélectifs. En effet, dans ces systèmes, l'apparition de structures a lieu *a posteriori*. Et, étant donné la diversité des répertoires du cerveau, il est extrêmement improbable que deux événements sélectifs donnés, même s'ils sont apparemment identiques, aient des conséquences identiques. Chaque individu est non seulement soumis, comme tous les systèmes matériels, au second principe de la

thermodynamique, mais sa perception et sa mémoire sont également soumises à un ensemble multistratifié d'événements sélectifs irréversibles. En fait, les systèmes sélectifs sont irréversibles par nature.

Cette « double exposition » de l'individu – d'une part, aux modifications du monde réel qui affectent les objets dépourvus d'intentionnalité et, d'autre part, aux modifications historiques individuelles survenant dans sa mémoire de sujet ayant une attitude intentionnelle – a d'importantes conséquences. En particulier, le flux de catégorisations dans un système sélectif conduisant à la mémoire et à la conscience modifie les relations de causalité habituelles décrites par les physiciens. Les individus, tout comme les choses, parcourent des trajectoires situées dans l'espace-temps à quatre dimensions. Mais, du fait que les êtres humains possèdent une intentionnalité, une mémoire et une conscience, ils sont capables de choisir des configurations apparaissant en un point donné de cette trajectoire et, en se fondant sur leurs histoires personnelles, de les soumettre à des plans en d'autres points de la trajectoire. Ils peuvent ensuite mettre ces plans à exécution, et modifier ainsi les relations de causalité entre objets d'une façon précise, conformément à la structure de leurs souvenirs. Tout se passe comme si des fragments d'espace-temps pouvaient glisser et être appliqués sur d'autres fragments. À ceci près, bien sûr, que cette transaction ne fait intervenir aucune bizarrerie physique, mais seulement la capacité de classer des choses par catégories, de mémoriser, et d'établir des plans d'après un modèle conceptuel. Des modifications historiques des chaînes causales d'une telle richesse ne pourraient jamais se produire dans aucune combinaison d'objets inanimés, dépourvus d'intentionnalité, car ces objets ne possèdent pas le type de mémoire adéquat. Il s'agit là, d'ailleurs, d'un important élément de distinction entre la biologie et la physique – une question que j'aborderai de manière plus détaillée au chapitre 20.

Dans certains systèmes mnésiques, des événements historiques singuliers survenant à une échelle donnée peuvent avoir des conséquences à une échelle complètement différente. Ainsi, par exemple, si le patrimoine génétique d'un lointain ancêtre a souffert des modifications à la suite des déplacements de cet ancêtre à travers un marécage (sous l'effet, disons, de fluctuations climatiques), la nouvelle séquence de nucléotides – s'il se trouvait qu'elle contribue à améliorer l'adaptation – pourrait avoir une influence sur les événements sélectifs et sur la fonction animale actuels. Et pourtant,

les lois physiques qui régissent les interactions *chimiques* survenant effectivement au niveau des éléments génétiques du patrimoine (les nucléotides) sont déterministes. Il n'en demeure pas moins qu'à elles seules, les lois déterministes au niveau chimique ne permettraient jamais d'expliquer les modifications génétiques régulièrement déclenchées, puis stabilisées sur de longues périodes, résultant d'événements sélectifs complexes survenus à l'échelle de l'animal tout entier dans des environnements singuliers.

Les phénomènes mnésiques survenant dans des cerveaux soumis à des phénomènes sélectifs sont du même acabit. Comme l'environnement en cours de catégorisation regorge de nouveautés, comme la sélection s'effectue *a posteriori* et comme elle se déroule sur des répertoires historiques hautement diversifiés, dans lequel des structures différentes peuvent produire des résultats identiques, le nombre de degrés de liberté est très grand. Nous pouvons donc conclure, sans risquer de nous tromper, que, dans un système conscient à plusieurs niveaux, le nombre de degrés de liberté est encore plus élevé. Ces constatations incitent à penser que les systèmes qui effectuent des catégorisations à la manière du cerveau présentent une indétermination au niveau macroscopique. De plus, il résulte de nos arguments précédents, concernant les effets de la mémoire sur la causalité, que la conscience autorise des « glissements temporels » lors de la planification, et que cela modifie le cours des événements.

Malgré ses succès en physique, en chimie et en biologie moléculaire, le réductionnisme n'en est pas moins absurde lorsqu'on l'applique exclusivement à la matière de l'esprit. Les mécanismes mentaux vont au-delà des principes newtoniens de causalité. Et les mécanismes mis en œuvre dans les mémoires d'ordre supérieur se situent au-delà des descriptions physiques de la succession temporelle. Enfin, l'existence d'individualités singulières au sein de la société constitue, dans une certaine mesure, un accident historique.

Ces conclusions nous amènent au problème classique du libre arbitre et à l'idée de « déterminisme mou », ou « compatibilisme », selon le terme dû à James Mill. Si ce que j'ai dit est exact, les êtres humains ont un certain degré de libre arbitre. Cette liberté n'est cependant pas totale ; elle est entravée par un certain nombre d'événements et de contraintes internes et externes. Un tel point de vue n'exclut pas l'influence de l'inconscient sur le comportement, et il ne sous-estime pas non plus le fait que de minuscules modifications biochimiques ou des événements précoces puissent influen-

cer le développement d'un individu de façon radicale. Mais il permet d'affirmer, en revanche, que le fort déterminisme psychologique postulé par Freud ne tient pas. Nous avons au moins la liberté que nous autorise notre grammaire.

Ces réflexions, et la relation existant entre notre modèle de la conscience et les valeurs dérivées de l'évolution, nous ramènent également à notre idée de la signification. La signification prend forme à partir de concepts qui dépendent de catégorisations fondées sur des valeurs. Elle se développe à mesure que l'histoire des sensations corporelles et des images mentales remémorées s'allonge. Ce mélange d'événements est propre à chaque individu et, dans une large mesure, imprévisible. Et, quand des compétences linguistiques et sémantiques apparaissent dans une société, et que des phrases contenant des métaphores sont associées à la pensée, la capacité de créer de nouveaux modèles du monde s'accroît vertigineusement. Mais il ne faut pas oublier que, en raison des liens qu'il entretient avec les valeurs et avec le concept de moi, ce système de significations n'est presque jamais indépendant des affects ; au contraire, il est soumis à une forte charge émotionnelle. Je ne tiens pas à aborder ici la question des émotions – les objets mentaux les plus complexes qui soient –, pas plus que celle de la pensée elle-même ; je les ai remises au chapitre suivant. Mais il me semble utile de les évoquer ici, par rapport au libre arbitre et à la signification. Comme l'ont souvent fait remarquer les philosophes et les psychologues, l'étendue de la liberté humaine est limitée par le fait que les individus sont incapables de séparer les conséquences de la pensée de celles des émotions.

Les êtres humains, créés à l'issue d'une séquence d'événements hautement improbable et soumis à des contraintes extrêmement strictes par leur histoire et leur morphologie, peuvent quand même faire preuve d'une extraordinaire liberté d'imagination. À l'évidence, ils ne sont pas du même ordre que les objets dépourvus d'intentionnalité. Ils sont capables de se référer au monde de façons très variées. Ils peuvent concevoir des plans, avoir des espérances et choisir d'influer de façon causale sur les événements du monde. Ils sont liés de nombreuses manières, accidentelles ou non, à leurs parents, à la société et au passé. Ils possèdent une identité consolidée par les émotions et la conscience d'ordre supérieur. Et ils ont quelque chose de tragique, dans la mesure où ils sont capables d'imaginer leur propre extinction.

On dit souvent que plusieurs épisodes de décentrage – à commen-

cer par la destruction des cosmologies plus anciennes, qui plaçaient l'humanité au centre de l'univers – ont fait subir aux êtres humains des pertes irréparables. Le premier de ces épisodes eut lieu, selon Freud, lorsque le géocentrisme fut remplacé par l'héliocentrisme. Le second se produisit lorsque Darwin mit en évidence l'origine des êtres humains, et le troisième lorsqu'il fut démontré que l'inconscient avait des effets puissants sur le comportement. Mais, bien avant Darwin et Freud, la vision newtonienne de l'univers avait déjà conduit à un fatalisme profond – à une vision qui mina les espérances sociales des penseurs du Siècle des Lumières. Cependant, nous pouvons à présent constater que, si les nouvelles idées concernant la fonction cérébrale et la conscience sont correctes, cette vision fataliste des choses n'est pas nécessairement justifiée. Le présent ne porte pas en lui la semence d'un futur fixement programmé, et il n'y a pas non plus de programme dans nos têtes. Les théories physiques modernes et les découvertes des neurosciences excluent non seulement les modèles mécanistes du monde, mais aussi les modèles mécanistes du cerveau.

Nous pouvons donc effectivement espérer que, lorsque des idées suffisamment générales, permettant d'effectuer la synthèse des découvertes provenant des neurosciences, seront émises, elles contribueront peut-être à fonder une seconde philosophie des Lumières. Et si cela se produit, cette philosophie s'appuiera principalement sur les neurosciences, pas sur la physique.

Le problème ne sera plus alors celui de l'existence des âmes, car il est clair que chaque individu est unique et qu'il n'est pas une machine. Le problème consistera à accepter que chaque esprit individuel est mortel. Mais, étant donné la conception laïque qui prévaut aujourd'hui, conception que nous avons héritée de la première philosophie des Lumières, comment préserver une morale tout en étant mortels ? Dans le cadre des modèles mécanistes actuels de l'esprit, ce problème prend des proportions majeures, car ces modèles permettent aisément de rejeter un être humain ou d'exploiter un individu comme s'il s'agissait d'une simple machine. Désormais, le mécanisme côtoie le fanatisme : les sociétés sont soit entre les mains d'hommes d'affaires puissants, mais spirituellement creux, soit, dans une moindre mesure, entre celles de fanatiques qui se trouvent sous l'emprise de mythes et de sentiments totalement non scientifiques. Peut-être que, lorsque nous aurons compris et adopté une vision scientifique de la façon dont notre esprit est apparu dans le

monde, notre conception de notre propre nature deviendra plus riche et les mythes que nous invoquerons plus indulgents.

Comment le fait de croire à une conception de la perception et de la conscience fondée sur le cerveau va-t-il affecter l'humanité ? Si nous acceptons l'idée que l'« esprit » de chaque individu est réellement incarné ; que cet esprit est précieux justement *parce* qu'il est mortel et que sa créativité est imprévisible ; que nous devons rester sceptiques quant à l'étendue possible de notre savoir ; qu'il est essentiel de comprendre comment se déroule le développement psychique chez les plus jeunes ; que l'imagination et la tolérance vont de pair ; que nous sommes tous frères et sœurs, du moins au niveau des valeurs fixées par l'évolution ; que bien que les problèmes moraux soient universels, leur résolution dans chaque cas particulier, si elle est possible, doit nécessairement prendre en compte l'histoire locale – si donc nous acceptons toutes ces idées, quelles conséquences cela entraîne-t-il ? Peut-on définir une morale convaincante en dépit du fait que l'esprit soit mortel ? Il s'agit là d'un des plus grands défis de notre époque.

Tant que les neurosciences n'auront pas atteint un certain degré de maturité, nous ne saurons pas relier clairement ces questions à notre histoire en tant qu'individus appartenant à une espèce dont l'évolution se poursuit toujours. Mais, quoi qu'il en soit, le réductionnisme idiot et le mécanisme simpliste sont hors course. Et il existe à présent une théorie de l'action fondée sur la notion de liberté humaine – justement ce qui manquait au Siècle des Lumières – qui semble être de plus en plus confirmée par les faits expérimentaux. Nous pouvons à présent examiner la relation existant entre ces faits et la pensée elle-même.

Chapitre 17

Pensées, jugements, émotions : des produits supérieurs

Il y a en nous plus savant que notre tête.
Arthur Schopenhauer

Comment se fait-il qu'un livre sur la question de l'esprit accorde si peu d'importance à la pensée, à la volonté et au jugement, ou encore aux sentiments, aux émotions et aux rêves ? En fait, c'est dû en partie à mes intentions initiales, qui étaient de décrire scientifiquement les bases nécessaires à la conscience et à la signification. J'ai tenté de le faire dans l'espoir qu'une fois que cette description aura été confirmée, on pourra lancer des recherches plus poussées et plus suffisantes en psychologie. Mais pour traiter l'un quelconque de ces produits supérieurs de l'activité mentale, il faudrait écrire un autre livre. Néanmoins, je voudrais faire ici quelques remarques sur la façon dont nos thèses sur l'esprit pourraient être reliées aux activités psychologiques.

Certains considèrent que la conscience et la pensée sont une seule et même chose. Je pense qu'une telle assimilation est trop simpliste, car la pensée comporte un certain nombre de composants supplémentaires acquis – un complexe d'images, d'intentions, de suppositions et de raisonnements logiques – et qu'elle constitue donc un mélange de différents niveaux d'activité mentale. Sous ses formes les plus élevées et les plus abstraites, il s'agit d'une compétence qui dépend des capacités symboliques de l'individu. À l'exception des capacités spatiales dont fait preuve la pensée des artistes plastiques, et des activités tonales et rythmiques qui accompagnent la pensée musicale, les activités supérieures de la pensée dépendent fortement à la fois du langage et de la logique, d'un dialogue interne entre celui qui pense et un autre « interlocuteur » dont le penseur peut d'ailleurs ne pas être conscient. C'est cela le « deux en un » auquel fait allusion Hannah Arendt dans son livre intitulé *La Vie de l'esprit*. Elle fait remarquer à ce propos qu'il existe une distinction en allemand entre la *Vernunft* – la raison ou pensée pure – et le

Verstand – la compréhension qui est directement liée aux processus cognitifs de la perception, des sentiments, et à d'autres du même type.

Je ne suis pas sûr que cette distinction soit utile du point de vue scientifique, mais elle sert à mettre l'accent sur le fait que la pensée peut aller extrêmement loin. Le penseur qui s'adonne à la pensée pure est à tel point plongé dans un état d'attention spécifique, lié au projet de penser, qu'il ou elle « s'abstrait » littéralement – il devient indifférent au temps qui passe, à l'espace, à sa propre personne et à ses perceptions. On peut dire que, à ces niveaux de signification et d'abstraction, « la pensée ne se trouve nulle part ». Mais il ne s'agit là que d'une métaphore permettant d'exprimer à quel point l'individu réussit à devenir insensible aux autres activités parallèles de l'esprit.

Quelle que soit la compétence exploitée par la pensée – qu'il s'agisse d'un talent pour les symboles logiques, mathématiques, linguistiques, spatiaux ou musicaux – il ne faut pas oublier que la pensée est mue par les processus jamesiens, qu'elle s'envole et qu'elle se pose, qu'elle est sujette à de grandes variations de l'attention et qu'en général, elle est alimentée par des processus métaphoriques et métonymiques. Ce n'est que lorsque les résultats de nombreux processus parallèles, fluctuants et temporels de perception, de formation de concepts, de mémoire, ainsi que de nombreux états d'attention, ont été « stockés » dans un objet symbolique – une séquence de propositions logiques, un livre, une œuvre d'art, une pièce de musique – que nous avons *l'impression* que la pensée est pure. Du fait que les pensées sont suscitées par d'autres pensées, par des images, et par un objectif mentalement conçu, nous avons l'impression qu'il existe un domaine de *Vernunft* – un lieu où le penseur (dans un état d'absorption profonde) ne se trouve nulle part et où le temps n'existe pas. De là au platonisme et à l'essentialisme, tous deux scientifiquement intenables, il n'y a qu'un pas.

Il est impossible de penser en l'absence d'une toile de fond consciente. Mais une théorie biologique de la conscience ne peut fournir qu'une condition nécessaire à la pensée, et non une condition suffisante. La pensée est une compétence que l'on construit à partir de l'expérience vécue, en entre-tissant les niveaux et les canaux parallèles de la vie perceptive et conceptuelle. Au bout du compte, il s'agit d'une compétence soumise aux contraintes des valeurs sociales et culturelles. L'acquisition de cette compétence exige plus que l'expérience des choses ; elle exige des interactions

sociales, affectives et linguistiques. Les pensées, les concepts et les croyances ne s'individualisent que par référence aux événements du monde extérieur, et par référence aux interactions sociales avec autrui – notamment à celles qui font appel à l'expérience linguistique.

Cela signifie que les données issues des neurosciences, aussi nombreuses soient-elles, ne permettront jamais, à elles seules, d'expliquer ce qu'est la pensée. Mais cette affirmation n'a rien de mystérieux ni de mystique : elle signifie simplement que l'explication issue des neurosciences est nécessaire, mais qu'elle n'est pas suffisante en tant qu'explication ultime. De même, bien qu'une description embryologique complète soit nécessaire pour expliquer pourquoi j'ai l'air d'un homme et que j'agis comme un homme, une telle description ne permettra jamais d'expliquer pourquoi je *suis* un homme. Seule l'adjonction d'une description en termes évolutifs, faisant intervenir des phénomènes historiques et la sélection naturelle, pourra fournir une explication suffisante.

Par conséquent, dans la pratique, toute tentative pour réduire la psychologie à la biologie finit nécessairement par échouer à un certain point. Étant donné que l'exercice de la pensée, en tant que compétence, dépend d'interactions sociales et culturelles, de conventions, de raisonnements logiques, et aussi de métaphores, les méthodes purement biologiques telles qu'on les connaît aujourd'hui sont insuffisantes. Cela est dû, en partie, au fait qu'à ses niveaux les plus élevés, la pensée est récursive et symbolique. En fait, comme chacun de nous est une source d'idiosyncrasies d'interprétation sémantique, et que la communication intersubjective (avec un interlocuteur réel ou imaginaire) est essentielle à la pensée, nous devons utiliser et étudier ces facultés pour elles-mêmes. Cependant, cette nécessité n'est pas incompatible avec notre idée selon laquelle la psychologie cognitive ne peut être correctement comprise qu'en présence d'une explication solide, fondée sur la biologie, de la conscience et des processus d'incarnation du sens.

Dans *Acts of Meaning*, Jerome Bruner donne des arguments très convaincants en faveur de l'idée que la construction du sens joue un rôle central en psychologie humaine. Il insiste sur le fait que le moi naît des interactions interpersonnelles qui ont lieu au sein de la culture sous l'influence de narratives, et il nous prie instamment d'utiliser des méthodes d'interprétation rigoureuses en psychologie sociale. Dans ce cas, nous sommes nos propres instruments scientifiques, et nous n'avons pas à être remplacés par des dispositifs

de mesure. L'objectif du présent ouvrage est justement de décrire les bases biologiques de la construction sémantique qui sous-tendent ce type de recherches en psychologie. Grâce à cette assise, nous pourrons voir comment la conscience, qui est fondée sur des systèmes de valeurs issus de l'évolution, et dont le moteur est le langage, conduit à l'extension et à la modification de ces systèmes au sein d'une culture.

Si cependant nous voulons déterminer quel est le moteur de la pensée, et aussi ce dont elle s'accompagne chez un individu donné, nous devons tout de même examiner l'état biologique de cet individu et explorer sa mémoire au niveau de la pensée pour voir comment elle induit, ou motive, d'autres pensées. Mais, par sa nature même, cette démarche est limitée. Nous devons donc également nous intéresser à l'étude des significations intrinsèques, des constructions mentales invariantes telles que les mathématiques, de l'invariance par rapport aux substitutions lexicales telle qu'on la trouve en logique et, plus généralement, à un ensemble de règles issues des interactions sociales et de l'expérience acquise. Il restera peut-être même un peu de place pour les philosophes, sans que pour autant nous leur accordions le privilège – erroné mais consacré par l'usage – d'utiliser la pensée comme *unique* moyen de comprendre l'origine de l'esprit.

De même que deux sciences différentes mais compatibles ne sont jamais totalement réductibles l'une à l'autre – l'une d'elles étant nécessaire, mais non suffisante, à l'autre – la description de la matière de l'esprit ne fait que fournir une base à l'analyse des questions relationnelles et symboliques. Mais, en construisant une telle description, on ne peut qu'être surpris par la nature multiple, parallèle et changeante des états conscients. La tâche des sciences cognitives consiste à découvrir comment interpréter les états correspondant à des modes symboliques de raisonnement, à des jugements et à la volonté, dans lesquels le sujet est plus directement conscient de son rapport avec le temps. Cependant, même si on réussissait à analyser ce type de questions, cela ne fournirait pas une explication adéquate de l'utilisation potentiellement illimitée des modes de raisonnement récursifs – induction, analogie et logique formelle. Et de toute façon, une telle analyse ne fournirait pas non plus une explication exhaustive des questions historiques.

Des régularités biologiques sous-tendent toutes ces activités. Ces régularités peuvent, et doivent, être étudiées. Mais tant que nous n'aurons pas construit des objets conscients capables de parler – ce

qui n'arrivera pas avant longtemps – les méthodes biologiques resteront trop maladroites pour permettre d'établir des corrélations entre le neuronal et la signification des pensées d'un « penseur pur » au cours d'un raisonnement. Néanmoins, nous pouvons quand même étudier les processus neuronaux de base qui sous-tendent ces actes, et nous pouvons le faire sans devenir des dualistes des propriétés. Mais, dans la pratique, il serait absurde, au nom de la pureté scientifique, de n'utiliser que des méthodes biologiques.

On pourrait en dire bien davantage, mais cela ne nous éclairerait pas plus sur ce qui nous intéresse, c'est-à-dire sur les bases biologiques de l'esprit. Il peut être utile, cependant, de faire quelques commentaires sur certaines questions apparentées, et en particulier sur celles qui concernent les sentiments et les émotions. Les sentiments font partie de l'état conscient ; ce sont les processus que nous associons à la notion de sensation dans la mesure où ils sont liés au moi. Ce ne sont pas des émotions cependant, car les émotions possèdent de fortes composantes cognitives, qui entremêlent les sentiments à la volonté et aux jugements d'une façon extraordinairement complexe. De fait, on peut dire que les émotions sont les états ou processus mentaux les plus complexes qui soient, dans la mesure où ils se mélangent avec tous les autres processus (d'une façon qui d'habitude est très spécifique, et qui dépend de l'émotion considérée). Et le fait qu'ils possèdent en outre des bases historiques et sociales n'est pas pour les simplifier.

Ce que les êtres humains conscients ont peut-être de plus extraordinaire, c'est leur art – leur capacité de transmettre des sentiments et des émotions, de manière symbolique et formelle, à travers des objets externes tels des poèmes, des tableaux ou des symphonies. La somme d'états conscients – conditionnés par des contraintes historiques et culturelles, par des apprentissages spécifiques et des compétences particulières – qui entre dans la composition d'une œuvre d'art ne peut être analysée par les méthodes scientifiques. Mais, encore une fois, une telle négation n'a rien de mystérieux, puisque nous ne pouvons comprendre ces objets et y réagir que par rapport à *nous-mêmes*, selon un mode social et symbolique. Aucune analyse externe objective, même si elle était possible, ne pourrait remplacer les réactions individuelles et les échanges intersubjectifs qui ont lieu au sein d'une tradition et d'une culture données. Suzanne Langer, dans son chef-d'œuvre *Mind : An Essay on Human Feeling*, analyse merveilleusement bien ces processus psychologiques.

Hormis la *Vernunft* et le *Verstand*, il existe en allemand un autre ensemble de mots, utilisés pour caractériser les connaissances humaines, qui établit lui aussi une distinction, comme cela a été mis en évidence pour la première fois par Wilhelm Dilthey. Le mot *Naturwissenschaften* fait référence à des connaissances ayant trait à ce que nous avions coutume d'appeler les sciences naturelles – la physique, la biologie, etc. Le mot *Geisteswissenschaften* fait, quant à lui, référence aux champs de la connaissance liés aux sciences sociales, à la culture, au raisonnement abstrait et à l'étude des événements historiques à partir des symboles et des sentiments. Cependant, cette distinction ne doit pas nous faire succomber à l'idéalisme de Hegel, qui fut parmi les plus éminents défenseurs du *Geist* (de l'Esprit), ni nous inciter à penser que la psychologie tombe hors du domaine de la biologie de l'évolution – c'est-à-dire qu'il existe un *Geist* d'un côté et une *Natur* de l'autre –, car cela conduit à une suite sans fin de complications superflues.

James, qui avait beaucoup réfléchi aux sujets abordés dans ce chapitre, disait ceci à propos de ce type de questions :

> La grande leçon que nous pouvons tirer de toutes les spéculations kantiennes et post-kantiennes est, me semble-t-il, une leçon de simplicité. Chez Kant, la complexité de pensée et d'expression constituait un handicap inné, exacerbé par l'académisme vieux jeu de la vie à Königsberg. Chez Hegel, il s'agissait plutôt d'une fièvre délirante. Ainsi, malheureusement, les raisins amers que ces pères de la philosophie ont mangés n'ont cessé de nous agacer les dents.

Étant donné ce que j'ai dit ici, je pense que la psychologie philosophique poursuivra sa propre route, à cette nuance près que, malgré les différences méthodologiques existant entre les *Geisteswissenschaften* et les *Naturwissenschaften*, la psychologie ne peut désormais se déclarer autonome par rapport à la biologie, et qu'elle doit toujours céder la priorité aux découvertes de la biologie.

Je me suis toujours demandé pourquoi il y avait un si grand nombre de sujets dans la liste de cours proposée par les universités. Pourquoi le savoir est-il si hétérogène ? Le point de vue présenté ici nous offre une explication possible. Étant donné les processus cérébraux parallèles et constructifs qui sous-tendent la conscience, étant donné les propriétés symboliques récursives du langage et, enfin, étant donné les bases historiques irréversibles de chaque réalisation symbolique et artistique au sein des sociétés et des

cultures, il ne peut exister de description complètement réductible des connaissances humaines. Néanmoins, les diverses sphères de la connaissance et les différents domaines peuvent être compatibles entre eux, et leurs bases, qui sont du ressort de l'évolution biologique et culturelle, peuvent être comprises. Les êtres humains, du moins en ce qui concerne leur étude de ces différents domaines, semblent avoir à peu près adopté la meilleure démarche possible.

Une situation plus poignante se crée lorsqu'un être humain souffre de perturbations neurales. Comme j'espère parvenir à le montrer au chapitre suivant, ces affections révèlent également l'immense éventail de réactions dont le système nerveux est capable et la complexité stratifiée dont il fait preuve.

Chapitre 18

Maladies mentales : le moi réintégré

Il est impossible de comprendre l'ensemble des perturbations neuropsychotiques sans les relier à des hypothèses claires sur les processus mentaux normaux.

Sigmund Freud à Wilhelm Fliess

Les maladies mentales ont toujours été considérées comme mystérieuses. Elles affectent l'« âme » de l'individu et, comme elles représentent une aberration par rapport à l'histoire et au comportement passés de cet individu, elles semblent étranges, si ce n'est à l'individu lui-même, du moins à ceux qui le connaissaient avant. Il est souvent difficile d'identifier les causes de ces maladies et, même lorsqu'on est convaincu qu'elles sont dues à des altérations de la fonction cérébrale, leurs symptômes ne sont pas aussi « directs » que ceux d'un grand nombre de perturbations neurologiques résultant également de dysfonctions cérébrales. Quelle est donc la différence ? Où se trouve la frontière entre les deux types de maladies ? En fait, il s'agit d'une différence subtile, mais presque toujours liée à des modifications de l'intentionnalité, de la conscience, des valeurs ou de la fonction symbolique. Une théorie de l'esprit telle que la nôtre dit clairement que toutes les maladies mentales sont dues à des modifications physiques.

Je n'ai pas du tout l'intention de traiter ce sujet fascinant et difficile. Mais si nous en négligions toutes les facettes – et elles sont nombreuses –, nous laisserions passer une occasion, fournie par la nature, de vérifier certains de nos modèles de l'esprit. Par conséquent, je vais quand même aborder ici quelques problèmes médicaux classiques sous l'angle qui nous intéresse dans ce livre. Il n'y a pas de meilleur moyen de mettre en évidence les mécanismes de contrôle qui affectent le cerveau et l'esprit à plusieurs niveaux. Tout d'abord, je compte m'interroger sur la nature des maladies mentales – s'agit-il ou non de maladies physiques ? Puis, je voudrais me demander pourquoi elles semblent différer des maladies neurologiques. Ensuite, je considérerai un certain nombre de maladies de la conscience – à la fois neurologiques et psychiatriques –, parce

qu'elles nous aident à mieux comprendre nos modèles. Enfin, je m'intéresserai aux maladies mentales en tant qu'adaptations, en tant que réintégrations du moi lorsque celui-ci est soumis à des conditions invalidantes. J'ai déjà abordé en partie ces questions dans *The Remembered Present*, mais pas tout à fait de la même manière. J'invite donc le lecteur à comparer les deux approches.

Freud était profondément intéressé par l'un des problèmes qui se trouvent au centre des questions que nous avons abordées au cours du présent ouvrage. En effet, au cours de ses recherches sur les névroses, il s'est aventuré dans les fourrés de l'intentionnalité. Pour Freud, les névroses étaient des altérations du comportement ou de l'affect qui laissaient intacte la capacité de « tester la réalité », mais qui perturbaient la fonction ou la satisfaction. Il ne pensait pas que ces perturbations étaient provoquées par des affections cérébrales. Il pensait plutôt qu'il s'agissait d'affections *fonctionnelles*, dues à des facteurs psychologiques, à des symbolismes à la fois conscients et inconscients. Les névroses étaient, selon lui, des maladies du *développement* psychologique. Il fonda sa théorie psychologique sur les résultats issus d'une méthode thérapeutique – la psychanalyse. Cette méthode fait intervenir des interactions symboliques entre un patient et un analyste formé à la méthode et disposant d'une théorie précise sur la façon dont se développe la personnalité humaine et se forme le moi de chaque individu. Le patient est incité à explorer, avec l'analyste, les mécanismes de défense et de refoulement en utilisant les techniques de l'association libre, de l'interprétation des rêves, et autres.

Bien que Freud ait initialement adopté un matérialisme éliminatoire très strict (précisément le genre de réductionnisme dont je déplorais l'existence au chapitre précédent), il se tourna par la suite vers une espèce de dualisme des propriétés. Ainsi, tout en restant d'un déterminisme et d'un matérialisme très stricts, il affirma que les névroses ne devaient cependant être considérées qu'en termes psychologiques. En revanche, les vues de Freud sur la psychose, maladie dans laquelle l'individu devient réellement incapable de tester la réalité et pour laquelle il est possible d'identifier des causes organiques, étaient plus équivoques. Certaines psychoses étaient – et sont toujours – considérées comme « fonctionnelles » : la schizophrénie, certaines formes de psychose maniaco-dépressive et la paranoïa, par exemple. Ces maladies ont des histoires causales, ou étiologiques, très différentes de celles des psychoses organiques ou des maladies dégénératives du cerveau qui peuvent aussi aboutir à

des psychoses. Ce type d'observations pose des problèmes à toute théorie du cerveau.

D'après la TSGN, ces problèmes apparaissent parce qu'il est extrêmement difficile de distinguer les divers niveaux de contrôle de la fonction cérébrale. Et le fait d'avoir affaire à une population de réponses synaptiques – vu leur diversité et leur individualité – fait apparaître une difficulté supplémentaire. Le *vrai* problème, cependant, ne réside pas dans la complexité, mais dans l'évaluation erronée des niveaux de causalité. Toutes les maladies psychiatriques, même celles dont l'origine est liée à des troubles de la communication individuelle et sociale, ont des causes physiques. Et, étant donné la multiplicité de niveaux présentée par les systèmes réentrants du cerveau qui régissent les états conscients et inconscients, il n'est pas surprenant que différentes causes de maladie aboutissent à des perturbations semblables, ou qui coïncident partiellement, de la structure des réponses. Tôt ou tard, toutes les aberrations se reflètent au niveau des synapses. Mais, en même temps, les signaux complexes et les interactions avec l'environnement sont liés de façon structurée à la mémoire et au comportement à tous les niveaux évoqués précédemment. Il est souvent plus utile de considérer les maladies mentales comme des perturbations de la catégorisation, de la mémoire, de la réentrée et de l'intégration plutôt que comme des perturbations de la capacité de « tester la réalité ».

Une façon de le visualiser consiste à adapter les diagrammes de la conscience donnés aux chapitres 11 et 12 à nos objectifs présents (figure 18-1). Si nous considérons, par exemple, que les facteurs qui détruisent ou qui affectent les neurones de certaines parties des ganglions de la base sont à l'origine de la maladie de Parkinson, ou que d'autres facteurs, affectant le cortex moteur, donnent lieu à une paralysie motrice, nous n'avons aucun mal à diagnostiquer ces affections comme neurologiques – c'est-à-dire, comme « non mentales ». Mais si la maladie concerne des interactions entre les circuits réentrants et hautement parallèles du cerveau, ou si elle altère les liens existant entre les systèmes de valeurs et ceux qui régissent le comportement (voir la figure 12-4), nous aurons plutôt tendance à diagnostiquer la maladie comme une altération de la fonction mentale. Or, dans un cas comme dans l'autre, des causes physiques suffisent à rendre compte des troubles.

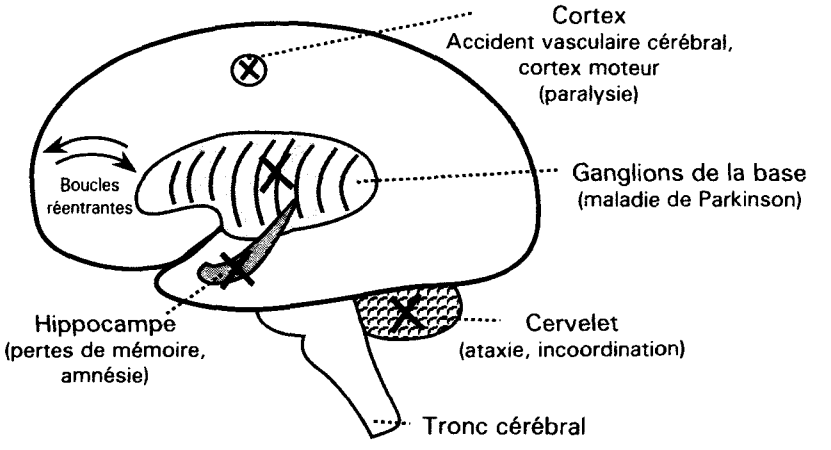
En fait, il est fructueux de considérer le problème de la maladie mentale sous l'angle des altérations des voies réentrantes et de la catégorisation. L'aliénation mentale et les maladies de la conscience

représentent des réarrangements et des adaptations face à des altérations des cartes réentrantes, des régions homéostatiques et des appendices corticaux responsables de la conscience perceptive, du fonctionnement symbolique conceptuel et des réactions émotionnelles.

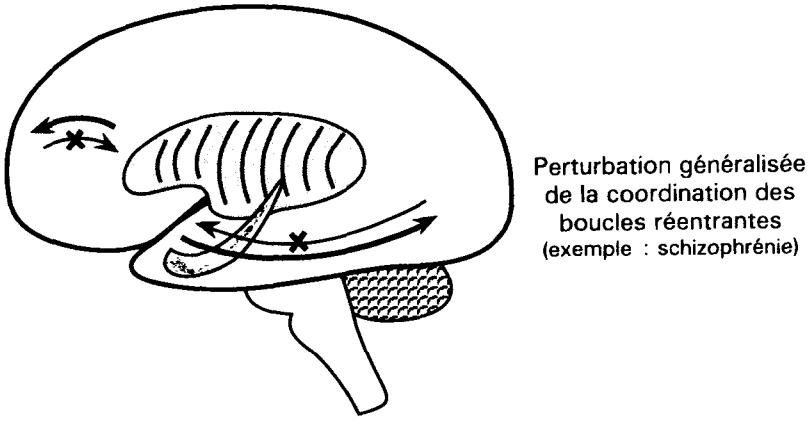
Quoi qu'il en soit, le fait que chacun des individus souffrant de ce type de troubles ait une histoire personnelle nous garantit que deux cas donnés ne seront pratiquement jamais pareils. La conscience d'ordre supérieur fait appel à des intégrations conceptuelles, sémantiques et sociales, qui toutes interviennent dans la construction du moi social. Un grand nombre de ces intégrations sont permises ou modulées par des populations de synapses particulières. Par conséquent, il n'est pas étonnant que l'on ait découvert que les médicaments capables de modifier la fonction synaptique sont extrêmement utiles pour traiter les troubles mentaux. Mais d'un autre côté, l'individualité de chaque patient conscient résulte de la configuration extrêmement complexe, et unique, de ses efficacités synaptiques. Par conséquent, l'utilisation exclusive de médicaments ne se substituera jamais à la communication verbale et affective qu'il est nécessaire d'établir avec le patient. Dans la plupart des cas, il faudra probablement allier un traitement médicamenteux et une psychothérapie.

Une théorie du cerveau qui considère la catégorisation, la mémoire et la formation des concepts en ces termes peut même servir dans des formulations purement psychothérapeutiques. Ainsi, une appréciation de la TSGN fut publiée il y a quelques années par un psychiatre du nom d'Arnold Modell, au cours de sa formulation d'une théorie du traitement psychanalytique. Dans son livre *Other Times, Other Realities*, Modell faisait appel à la notion de mémoire en tant que recatégorisation afin de réévaluer la nature des transactions entre le patient et le thérapeute. Il reprenait notamment le mot *Nachträglichkeit*, que Freud avait appliqué à l'idée que la mémoire est retranscrite à l'issue d'expériences ultérieures. Modell fait remarquer à ce propos que l'ego est une structure engagée dans le traitement et la réorganisation du temps, ce qu'il associe à l'idée de mémoire en tant que recatégorisation (voir les chapitres 10 et 16).

Dans sa critique des concepts associés à la relation de transfert dans le cadre psychanalytique, Modell suggère que l'adoption d'une vision sélectionniste de la fonction cérébrale fournit une interprétation alternative de la compulsion de répétition décrite par Freud. Selon lui, le fait que les catégories se recréent dans la mémoire



MALADIES NEUROLOGIQUES



MALADIES PSYCHIATRIQUES

FIGURE 18-1

Maladies du système nerveux et maladies de l'esprit. Le terme « maladie neurologique » fait référence aux perturbations de la vue, du mouvement, etc., qui résultent d'altérations des régions cérébrales concernées par ces fonctions. Le terme « maladie psychiatrique » fait référence à des perturbations de la catégorisation, de l'activité mentale, des sensations, etc., dans lesquelles les réponses deviennent aberrantes sur le plan symbolique ou dans lesquelles la capacité de « tester la réalité » est compromise. Ces maladies sont dues à des modifications fonctionnelles survenant à de multiples niveaux, des synapses aux boucles réentrantes. Les deux catégories de maladies sont d'origine physique, et elles se recouvrent partiellement. Les maladies psychiatriques affectent plus largement la catégorisation, la mémoire et les processus symboliques via des boucles réentrantes.

constitue un principe biologique de base qui, dans certaines circonstances, prévaut sur le principe de plaisir de Freud. Modell suggère que « la compulsion de répéter correspond à la recherche compulsive d'une identité perceptive entre les objets présents et passés ». Ce faisant, le patient est conscient, à des degrés variables, de la relation existant entre le moi et le temps, une question que nous avons déjà évoquée au chapitre 16. Modell fait également remarquer que, au cours du processus psychanalytique, la pensée métaphorique est monnaie courante dans l'esprit. Enfin, il suggère que la vision de la mémoire proposée par la TSGN, qui remplace la vision freudienne de l'existence de traces de mémoire fixes, permet de considérer le cadre thérapeutique sous un jour nouveau. Selon cette vision, le cadre thérapeutique est destiné « à accentuer les multiples niveaux de réalité, ce qui à son tour fait augmenter le potentiel d'anciennes perceptions et de retranscriptions de nouvelles perceptions ».

Un autre psychiatre – Edward Hundert – a également utilisé la TSGN comme pierre de touche pour relier la psychiatrie, la philosophie et les neurosciences. Selon lui, il faut que ces trois disciplines soient reliées à travers une théorie de ce type afin de garantir que les individus soient considérés sous un angle suffisamment large. Je conseille vivement aux lecteurs intéressés de comparer les remarques de Hundert et les notions que j'ai présentées au chapitre précédent et tout au long de ce livre.

Bien que je sois loin d'être une autorité en matière de théorie psychanalytique, je pense qu'il est important que des tentatives soient en cours qui visent à relier la théorie psychanalytique à une théorie du cerveau physiquement fondée qui s'intéresse aux problèmes de catégorisation. Quant aux tentatives visant à relier des effets pharmacologiques à des théories du cerveau, elles sont un peu plus nombreuses. Nous avons besoin de ces deux types de recherches si nous voulons bâtir une psychiatrie solidement ancrée dans la biologie.

Voilà pour les remarques d'ordre général. Je vais à présent me tourner vers un ensemble spécifique de maladies neurologiques et vers une maladie psychiatrique. J'y tiens, parce que ces deux types de maladies permettent de mieux comprendre mes propositions concernant l'importance qu'ont les circuits réentrants pour la conscience. Les troubles neurologiques que je vais aborder constituent un ensemble de syndromes affectant ce que l'on appelle la mémoire implicite. Quant à la maladie psychiatrique, il s'agit de la schizo-

phrénie – la plus tarabiscotée, la plus variable et la plus mystérieuse de toutes les psychoses.

Commençons donc par les syndromes neuropsychologiques (ici, le lecteur doit me permettre d'utiliser un peu de jargon médical ; je traduirai au fur et à mesure). Daniel Schacter et ses collègues ont étudié un ensemble de dissociations survenant entre l'accès explicite, conscient, au savoir, et la capacité implicite d'accomplir une tâche. Ce type de dissociations – un savoir implicite presque normal accompagné d'un savoir explicite gravement perturbé – a été observé chez des patients atteints de maladies diverses. Mais, avant de les énumérer et d'en décrire quelques cas, il est bon d'en donner un exemple concret : celui de l'agnosie visuelle. Les patients souffrant d'agnosie visuelle discernent des stimulus visuels dans l'espace bien qu'ils soient incapables de voir la partie du champ visuel dans laquelle ces stimulus sont présentés. Ces patients accomplissent correctement des tâches de discrimination visuelle alors même qu'ils ne se rendent pas perceptivement compte – c'est-à-dire qu'ils sont perceptivement inconscients – du fait qu'ils en sont capables.

De telles dissociations ont été observées chez des patients souffrant d'amnésie, de dyslexie (l'incapacité de lire certains textes), d'aphasie (l'incapacité d'exprimer, ou produire, des paroles intelligibles), de prosopagnosie (l'incapacité de reconnaître des visages), d'héminégligence (l'incapacité de faire attention à ce qui se passe dans la moitié de l'espace qui, du point de vue de l'individu, se trouve du côté opposé à l'hémisphère cérébral atteint) et l'anosognosie (le fait qu'un patient soit incapable de se rendre compte, ou même qu'il nie, l'existence de défauts neurologiques flagrants même lorsqu'on lui en donne des preuves directes).

Il s'agit là d'une liste extraordinaire. De fait, au cours de l'accomplissement des tâches, les amnésiques se montrent capables d'apprendre à réagir à certains stimulus – ce qui indique qu'ils en ont pris spécifiquement connaissance – alors même qu'ils nient être conscients de leur existence. Les prosopagnosiques se montrent capables de reconnaître implicitement des visages dont on sait qu'ils les ont déjà vus avant le début de leur maladie, mais dont ils ne sont à présent plus conscients. Dans certaines circonstances, les dyslexiques sont capables de lire des textes sans s'apercevoir qu'ils en sont capables. Et chez les patients atteints d'héminégligence, l'information présentée du côté négligé affecte l'accomplissement de certaines tâches sans que ces patients en soient conscients.

Les lésions sous-jacentes à ces affections sont toutes différentes, et aucun de ces cas ne s'accompagne d'une modification globale de la conscience. En fait, les patients se comportent de façon tout à fait normale en dehors des domaines présentant le défaut ; en d'autres termes, le défaut présent dans chaque cas est spécifique d'un domaine donné. Rien ne semble indiquer que des troubles du langage soient responsables des réponses dissociatives des patients. Et surtout, rien ne semble indiquer que les patients atteints de ces syndromes souffrent de névroses ou de psychoses.

Ces maladies sont des maladies de la conscience. Elles peuvent s'expliquer en supposant que ce qui a été affecté dans chacune d'elles est une boucle réentrante particulière reliant la mémoire des valeurs-catégories aux couples de classification chargés de la catégorisation perceptive (figure 18-1 ; voir également la figure 12-4). Remarquez que cette interférence n'affectera pas, en général, les *autres* voies de communication avec la mémoire des valeurs-catégories qui interviennent dans l'accomplissement d'une tâche spécifique. Il en résulte donc une délétion spécifique d'un domaine donné dans la « scène » consciente, mais *pas* dans le répertoire des capacités de l'individu relativement à l'accomplissement de la tâche. On peut se demander ce qu'il resterait de la conscience primaire si ces voies réentrantes étaient coupées une à une, modalité après modalité. Bien sûr, il est peu probable que cela se produise, sauf en cas de lésions massives et, pour des raisons éthiques, il est difficile d'imaginer que l'on puisse réaliser une telle expérience, même si l'on disposait de moyens moins traumatiques pour y parvenir.

J'ai voulu décrire ces troubles *neurologiques* pour montrer que, tout comme les maladies mentales proprement dites, ils peuvent également être *mentaux* – c'est-à-dire, ils peuvent affecter l'intentionnalité. Et, pour illustrer ces maladies de façon encore plus frappante, je vais clore ce chapitre en décrivant un cas de dissociation franchement mystérieux, à savoir celui de l'anosognosie. Mais avant cela, tournons-nous un instant vers la plus flagrante, la plus polymorphe et la plus mystérieuse des psychoses : la schizophrénie. La schizophrénie se caractérise par des mélanges de symptômes bizarres. Entre autres, les schizophrènes entendent des voix, se croient sous le contrôle de forces extra-terrestres, ont des délires de persécution et d'influence, éprouvent un barrage de la pensée, un écho de la pensée, croient qu'on devine et qu'on répète leurs pensées, ou encore qu'on les leur vole. Dans les stades aigus de ce syndrome, les symptômes s'accompagnent d'un état de rêverie

et d'un léger obscurcissement de la conscience, ainsi que d'un état de perplexité. Les patients se sentent submergés par des signaux qui n'ont de sens que par fragments et au sein d'îlots de lucidité. Certains voient leur affect s'émousser. Certains ont beaucoup de mal à se concentrer, évaluent incorrectement les signaux perceptifs, ont des hallucinations visuelles, ou sont incapables de discerner des formes gestaltistes. D'autres manquent de jugement, réagissent lentement, ou présentent des persévérations. Dans les formes extrêmes de la maladie, les patients entrent parfois en catatonie : ils ne réagissent plus, tout en maintenant des postures bizarres pendant de longues périodes. Lorsqu'ils sortent de ces épisodes aigus, certains patients se souviennent de ce qui s'est dit ou fait lorsqu'ils étaient en catatonie, tandis que d'autres n'ont que peu de souvenirs de cette expérience aiguë.

La schizophrénie est une maladie de la conscience aux formes variées, qui affecte la perception, la pensée et les sensations. Voir un patient qui souffre de cette affection aux prises avec sa constellation personnelle de symptômes constitue une expérience à la fois émouvante et pénible. Il n'a pas été facile de rendre compte de la diversité des symptômes, des caractéristiques individuelles et des bizarres particularités de cette maladie.

J'ai déjà suggéré que la schizophrénie pourrait être une affection généralisée de la réentrée (voir le bas de la figure 18-1). Mon hypothèse est que des perturbations survenant dans la production de plusieurs neuromédiateurs, ou dans la réponse à plusieurs d'entre eux, pourraient entraîner une mise hors d'état généralisée des communications entre cartes réentrantes. Le fait que la cartographie ne s'effectue pas de manière adéquate, ou que les différentes cartes ne soient pas correctement synchronisées, pourrait faire, chez un patient donné, que l'imagerie prédomine sur les entrées perceptives, ou encore que les différentes modalités ne soient plus correctement coordonnées. Cela pourrait donc conduire à des hallucinations ou à l'incapacité de coordonner les signaux provenant du monde réel. Chez un autre patient, la même affection pourrait entraîner des perturbations dans les connexions réentrantes reliant différents systèmes conceptuels aux organes de la succession. Et chez d'autres encore, il pourrait se produire des perturbations au niveau des liaisons réentrantes entre les aires ayant trait au lexique, les centres conceptuels et ceux responsables de l'imagerie mentale.

Le fait que les réactions varient d'un patient schizophrène à un autre peut s'expliquer par des facteurs historiques et personnels

faisant intervenir la variabilité des répertoires. Chaque patient présente une configuration unique de réponses, et ses réactions aux perturbations des divers sites réentrants sont différentes. Cependant, je ne veux pas dire par là que tous les neurones restent normaux dans la schizophrénie (ce dont je doute fort) ; je veux simplement dire que le *principal* défaut psychologique résulte d'une faille dans les cartographies réentrantes. Mais cette faille pourrait être due à n'importe lequel des facteurs susceptibles d'altérer les cartes elles-mêmes ou leurs connexions, et en particulier à la dysfonction neuronale ou à la mort des neurones eux-mêmes.

Il n'est pas difficile d'imaginer ce qu'un patient qui possède encore une conscience d'ordre supérieur ferait, s'il était atteint par cette maladie, pour tenter d'adapter son moi à ce qu'il ou elle perçoit. Et il est clair que ce comportement semblerait anormal d'après les critères normatifs et physiologiques. Mais il est néanmoins vraisemblable que la réaction globale du patient soit quand même une tentative d'adaptation, de réintégration. Elle peut ne pas être la meilleure stratégie d'adaptation possible, elle peut être destructrice pour le patient et pour les autres ; mais la question n'est pas là. Ce qui importe, c'est que l'esprit d'un patient schizophrène, vu sous l'angle de la TSGN étendue, possède une cohérence interne, surtout pour qui connaît l'histoire personnelle de ce patient. Mais hélas, la « cohérence » n'existe que du point de vue des prédictions de la théorie, et non du point de vue des valeurs sociales ou affectives spécifiques d'une société donnée.

Nous avons trop facilement tendance à sous-estimer les facultés des psychotiques, tout comme nous avons facilement tendance à interpréter incorrectement les excentricités apparentes des individus normaux. Je me souviens avoir entendu ou lu quelque part – je ne sais plus où – l'histoire d'un homme qui vivait à Paris sous l'occupation nazie et qui, se sachant sur le point d'être arrêté par la Gestapo, décida que la cachette la plus sûre était l'hôpital psychiatrique. Il cultiva donc l'art de se comporter comme un fou et réussit à se faire enfermer à la suite d'un épisode « hallucinatoire » survenu dans la rue. En adoptant, de temps en temps, des comportements délibérément bizarres devant les médecins et les autres malades, l'homme put ainsi vivre sans crainte pendant pas mal de temps. Mais un jour, deux sinistres individus vêtus de longs manteaux de cuir noir se présentèrent devant la porte de sa chambre accompagnés par le surveillant-chef. Convaincu qu'il s'agissait de membres de la Gestapo, l'homme se leva d'un bond, adopta une posture bizarre,

roula les yeux et commença à émettre d'étranges glapissements. C'est alors que l'homme qui occupait le lit voisin du sien, et qui passait le plus clair de son temps dans une espèce de transe, ouvrit les yeux et lui dit fermement : « Taisez-vous, simulateur [1]. »

En ce qui concerne la « simulation » normale (mais pas la vraie imposture), cela me rappelle mon cher ami Lars Onsager, aujourd'hui décédé – l'un des spécialistes de physico-chimie les plus remarquables de notre temps, et en fait le scientifique le plus extraordinairement doué que j'aie jamais rencontré. Lorsqu'il assistait à une conférence plutôt ésotérique, il laissait tomber la tête de côté, comme s'il dormait, cinq minutes après s'être assis. Mais si jamais le conférencier venait à se tromper dans une équation, il pouvait très bien arriver que Lars se lève, se dirige calmement vers le tableau, efface l'erreur, la corrige, lance un sourire, et retourne à sa place pour s'endormir à nouveau. Un jour, je lui demandai : « Lars, quand les gens te posent des questions profondes, pourquoi est-ce que tu souris, que tu glousses, que tu hoches la tête et que tu dis des choses incompréhensibles ? » Lars prit un air grave, presque austère, et me dit : « Je suis paresseux. Paresseux ! » Je m'écriai alors : « Je ne comprends pas. » Ce à quoi il répondit : « J'ai envie de répondre à *mes* questions, pas aux *leurs*. »

J'ai délibérément choisi deux exemples extrêmes, issus respectivement du « camp neurologique » et du « camp psychiatrique » – la dissociation implicite-explicite et la schizophrénie. Tous deux sont des exemples de maladies de la conscience. Et nous devons nous méfier du cartésianisme lorsque nous analysons des patients atteints de l'un ou l'autre de ces types de maladies. Pour vous en convaincre, je vais décrire un cas rapporté par le neurologue italien Eduardo Bisiach. Ce cas constitue, pour la théorie du cerveau, un défi aussi significatif que le fut celui d'Anna O. pour la théorie de Freud, au tout début de la psychanalyse.

Le patient de Bisiach était atteint d'anosognosie ; autrement dit, il niait l'existence chez lui d'un certain défaut neurologique, même lorsqu'on lui présentait des preuves directes de cette existence. Dans ce cas particulier, le syndrome s'accompagnait d'une hémiplégie gauche (une paralysie de la moitié gauche du corps) et d'une hémianopsie gauche (l'incapacité de voir le champ visuel gauche). Le syndrome avait été provoqué par un accident vasculaire, survenu de façon soudaine, qui avait affecté les aires temporelles, pariétales

1. En français dans le texte (N.d.T.).

et occipitales du cerveau. (En gros, ces aires interviennent dans l'accomplissement de tâches visuelles et confèrent également certaines capacités motrices.)

Le patient était intelligent, il avait de bonnes réactions et ne semblait pas perturbé du point de vue émotionnel. Il ne présentait aucun trouble de la parole. Mais il était anosognosique par rapport à ces déficiences à la fois visuelles et motrices. Lorsqu'on l'interrogeait à propos de tâches concernant le côté gauche qu'il n'avait pas accomplies, il affirmait les avoir accomplies. (Évidemment, la paralysie de son côté gauche rendait leur accomplissement impossible.) L'examinateur prenait alors la main gauche – paralysée – du patient dans les siennes et la plaçait dans le champ visuel droit du patient, de sorte que celui-ci puisse la voir. Mais lorsque l'examinateur demandait au patient à qui appartenait cette main, le patient affirmait qu'elle appartenait à l'examinateur. Et quand on lui demandait comment il expliquait le fait que l'examinateur ait trois mains, il répondait, avec une logique imparable : « La main se trouve toujours à l'extrémité d'un bras. Et, étant donné que vous avez trois bras, il s'ensuit que vous avez trois mains. »

Si nous partons du fait que ce patient n'était ni névrosé, ni psychotique, et qu'il n'avait aucun trouble du langage, nous sommes obligés d'admettre l'extraordinaire conclusion suivante : la conscience fondée sur le langage peut être altérée suite à la suppression, chez un individu, des sources cérébrales de signaux non verbaux. Et ce qu'il y a de frappant, c'est que ce patient avait réintégré l'ensemble de son interprétation sémantique de la réalité sans être le moins du monde affecté du point de vue émotionnel. Il avait subi une totale réorganisation et une complète réintégration *conceptuelles.*

Il serait logique de penser que si l'image de son propre corps et la capacité d'effectuer des catégorisations avaient été « stockées » dans la mémoire du patient, l'opposition existant entre cette mémoire stockée et son état perceptif et moteur ultérieur l'aurait conduit soit à une description cohérente, « réaliste », de l'état critique dans lequel il se trouvait effectivement, soit à des troubles émotionnels profonds. Mais si notre théorie est correcte, cela signifie que le patient ne possède pas une mémoire figée de ce type. Et, comme il est incapable d'accéder à certaines parties de son espace personnel, il a subi une réintégration conceptuelle et sémantique qui non seulement reflète cette incapacité mais qui aussi, dans une certaine mesure, construit autour d'elle une vision adaptative.

Ce cas montre clairement que des modifications de l'intentionnalité peuvent accompagner des maladies neurologiques. Mais en même temps il remet profondément en cause nos idées sur l'esprit. Comment le moi se réintègre-t-il ? À ce propos, je voudrais également évoquer les étonnantes réponses d'un individu souffrant d'un *split-brain* [2] et dont l'hémisphère gauche du cerveau semble appartenir à un individu plus ou moins normal doté d'une conscience d'ordre supérieur. Dans certains cas, son hémisphère droit répond à des mots, présentés visuellement, par des mots épelés de la main gauche à l'aide des lettres d'un jeu de Scrabble. Parfois, les réponses concordent avec celles de l'hémisphère gauche, et parfois non. Nous sommes donc obligés d'envisager la possibilité que la conscience primaire soit en train de se manifester par l'intermédiaire de l'hémisphère droit de son cerveau. Mais malheureusement, il est impossible d'effectuer une analyse complète en raison des interférences dues à des facteurs tels que l'apprentissage passé ou la connectivité qui existait auparavant entre les deux hémisphères cérébraux. Quoi qu'il en soit, la *personne* qui s'exprime à travers la conscience d'ordre supérieur siège dans l'hémisphère gauche et interprète tous les événements de façon adaptative.

Ces observations posent des problèmes intimement liés à ceux qui furent étudiés par Freud. Qu'est-ce qui régit la disposition des réponses conscientes et inconscientes de façon à permettre l'unification de la personnalité ? Quel est l'appareillage minimal nécessaire à l'apparition de la conscience d'ordre supérieur ? Qu'est-ce qui préside au déroulement de la réintégration lorsque l'individu est atteint par la maladie ?

L'étude de la maladie mentale à tous les niveaux est évidemment aussi essentielle pour comprendre le fonctionnement du cerveau que pour comprendre ce que veut dire être un individu au sein d'une société. Étant donné l'importance culturelle qu'elle revêt pour nous, l'étude de la maladie mentale est évidemment importante du point de vue pratique. Cependant, de telles études sont extrêmement complexes, et il est peu probable que la psychiatrie à elle seule suffise à dévoiler les mécanismes normaux et anormaux du cerveau. Cela exige la participation de nombreuses disciplines. Une des formes d'étude les plus éloignées de la psychiatrie consiste à tenter de fabriquer des objets ayant des fonctions psychologiques et faisant preuve d'intentionnalité. Si ces efforts aboutissent, ils nous aideront

2. Individu dont les deux hémisphères cérébraux ne sont plus connectés (N.d.T.).

considérablement à comprendre quelle est notre place dans la nature – que nous soyons bien portants ou malades. Par conséquent, je vais à présent quitter l'étude de l'intentionnalité chez l'être humain pour m'intéresser à la possibilité de créer des objets doués d'intentionnalité. Il s'agit là d'une perspective passionnante qui, quelles que soient ses limitations, revêt une énorme importance à la fois pratique et théorique.

Chapitre 19

Peut-on construire un objet conscient ?

> *On voit qu'il n'y [...] a qu'une [substance] dans l'Univers, et que l'Homme est la plus parfaite. Il est au Singe, aux Animaux les plus spirituels, ce que le pendule planétaire de Huygens est à une Montre de Julien le Roi. S'il a fallu plus d'instrumens, plus de Roüages, plus de ressorts pour marquer les mouvements des Planètes, que pour marquer les Heures, ou les repeter ; s'il a fallu plus d'art à Vaucanson pour faire son* Fluteur, *que pour son* Canard, *il eût dû en employer encore davantage pour faire un* Parleur *; Machine qui ne peut plus être regardée comme impossible, surtout entre les mains d'un nouveau Prométhée.*
>
> Julien Offray de la Mettrie

Ainsi placé – à la suite de plusieurs chapitres au cours desquels nous avons abordé diverses questions concernant l'être humain – le titre de celui-ci peut paraître incongru. Mais effectivement, j'ai maintenant l'intention de me demander si le fait de savoir comment fonctionne le cerveau nous permettra ou non de construire des objets doués d'intentionnalité. Je voudrais également suggérer que la seule façon de parvenir à intégrer efficacement nos connaissances sur le cerveau, étant donné la multiplicité de ses niveaux, consiste à synthétiser des objets. Pour y parvenir, nous avons besoin des ordinateurs les plus élaborés possible. Toutefois, les tentatives de construction d'objets « conscients », bien que peu nombreuses, ne datent pas d'aujourd'hui, comme en témoignent l'épigraphe et la figure 19-1.

Mais n'est-il pas en fait contradictoire de suggérer qu'il faut utiliser des ordinateurs dans un livre qui affirme que le cerveau n'est *pas* un ordinateur ? Pour répondre à cette question, je dois en dire un peu plus sur les ordinateurs. Ensuite, je m'interrogerai sur la possibilité de construire plusieurs types d'objets : une machine perceptrice, un objet doué de conscience primaire et enfin, un objet possédant une conscience d'ordre supérieur. Si la réponse à l'une quelconque de ces questions s'avère positive, alors nous *devrons* en considérer les implications morales, comme

c'est le cas pour l'application de n'importe quelle découverte scientifique.

Les ordinateurs sont des machines logiques ; en principe, ils sont capables d'exécuter n'importe quelle procédure effective, spécifiée sans équivoque par un ensemble d'instructions et produisant un résultat unique pour un problème donné. J'ai déjà dit plus haut que le cerveau n'est pas un ordinateur et que le monde n'est pas spécifié de façon suffisamment univoque pour être susceptible d'agir comme un ensemble d'instructions. Cependant, les ordinateurs peuvent être utilisés pour *simuler* des parties du cerveau et même pour aider à construire des machines perceptrices fondées sur la sélection plutôt que sur des instructions.

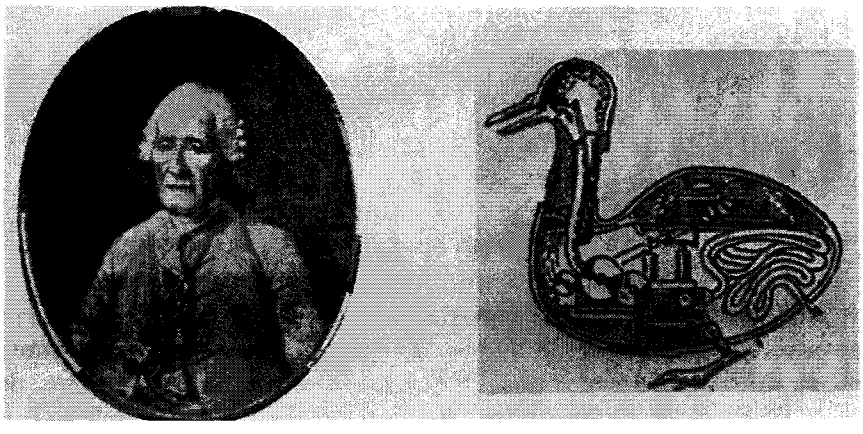


FIGURE 19-1

Jacques de Vaucanson (1709-1782), célèbre constructeur d'objets imitant des comportements, figure ici aux côtés de son canard. L'engin faisait coin-coin, se dandinait et possédait une « fonction intestinale ».

Pour lever cette apparente contradiction, il me suffit de préciser en quoi consiste une simulation. Pour faire une simulation, on écrit un programme qui spécifie les propriétés structurelles requises et les principes opératoires de l'entité que l'on veut simuler. Le programme est bâti de telle sorte que, lorsqu'il sera exécuté, certaines parties de l'entité que l'on simule dans son ensemble assureront leurs fonctions adéquates. Si, par exemple, je souhaite simuler un Boeing 747 entrant dans une zone de turbulences atmosphériques, je dois introduire dans le programme les caractéristiques de la forme de l'avion ainsi que les principes qui lui permettent de voler – ses propriétés aérodynamiques, la portance pour un poids donné, etc.

Et, si le programme est correctement conçu, je pourrai « faire voler » l'avion aussi bien dans des conditions de beau temps que dans des conditions turbulentes. Je pourrai ainsi, par exemple, voir si l'avion est susceptible de décrocher dans certaines conditions, ou si les ailes sont susceptibles d'être arrachées sous l'effet de vibrations trop violentes. S'il réussit, cet exercice aura été moins onéreux et fournira davantage d'informations que les essais d'une maquette réelle de l'avion dans une soufflerie.

Est-il possible de simuler un système sélectif ? La réponse doit être divisée en deux. Si je considère un animal *particulier*, résultat d'une sélection au cours de l'évolution et du développement, puisque je connais déjà sa structure et les principes qui régissent ses processus sélectifs, je peux simuler la structure de cet animal dans un ordinateur. Mais les systèmes qui sont en train de subir une sélection possèdent deux parties : un animal ou organe, et un environnement, ou monde (voir le chapitre 8). Et le système sur lequel s'exerce la sélection ne reçoit aucune instruction provenant des événements du monde. De plus, les événements qui surviennent dans l'environnement, ou monde, sont imprévisibles. Comment faire, alors, pour simuler des événements et leurs effets sur la sélection ? On peut procéder comme suit :

1. Simuler l'organe ou l'animal comme décrit ci-dessus, en tenant compte du fait que, en tant que système sélectif, il comporte un générateur de diversité – mutations, modifications des circuits neuronaux ou modifications synaptiques imprévisibles.
2. Indépendamment, simuler un monde ou environnement soumis aux contraintes créées par des principes physiques connus, mais tout en permettant à des événements imprévus de se produire.
3. Permettre à l'organe ou animal simulé d'interagir avec le monde simulé ou avec le monde réel en l'absence de tout transfert préalable d'information, de sorte que la sélection puisse avoir lieu.
4. Voir ce qui arrive.

Nous avons ainsi créé une situation dans laquelle des variations imprévisibles surviennent dans deux systèmes ou domaines distincts. De plus, si nous permettons à ces domaines d'interagir au cours de la simulation, il n'y a aucun moyen de savoir à l'avance ce qui va arriver (si ce n'est en effectuant la même simulation sur un autre ordinateur). Par conséquent, la procédure effective la plus simple qui permette de connaître l'issue de la simulation n'est autre que la simulation elle-même. En d'autres termes, le programme utilisé pour chaque domaine spécifie *bien* des contraintes et des

structures, mais il *ne peut pas* spécifier la façon précise dont elles interagiront dans des conditions de variabilité et de sélection.

La variabilité est introduite dans la simulation par l'intermédiaire d'un outil technique appelé générateur de nombres pseudo-aléatoires. Il s'agit d'une formule qui permet de simuler la production de nombres au hasard ! Elle n'est d'ailleurs pas parfaite ; si nous voulions saisir parfaitement le hasard, il nous faudrait, par exemple, brancher une source radioactive émettant des rayonnements alpha sur un compteur, *puis* relier le compteur à un ordinateur. Mais en fait, tout indique qu'il suffit d'utiliser deux générateurs pseudo-aléatoires différents – l'un pour la variabilité interne de l'animal et l'autre pour la variabilité de l'environnement – et de laisser la sélection opérer, pour parvenir à éviter d'introduire un biais prévisible dans le système. Après tout, ces deux systèmes n'ont aucun moyen de « savoir » quelles sont les variations qui vont de pair et, en outre, nous pouvons toujours décider de remplacer constamment les générateurs aléatoires.

Dans ces conditions, nous ne pouvons spécifier aucune procédure effective *donnant les conséquences de la sélection* qui soit indépendante de notre choix de générateurs de nombres pseudo-aléatoires. Et dans cette mesure, cela n'a pas de sens de dire que le système et ses résultats futurs se comportent *en bloc* comme un ordinateur (ou une machine de Turing). Bien sûr, tout résultat passé particulier de la sélection, *une fois connu*, pourra être ainsi spécifié. Mais, grâce à la procédure décrite ci-dessus, nous avons ajouté à notre ordinateur simulateur des caractéristiques qui le transforment en quelque chose qui ne se comporte plus strictement comme un ordinateur. Cela tient au fait que nous avons choisi des méthodes de génération de nombres aléatoires qui ne sont pas couplées de manière prévisible aux séquences d'événements survenant dans les systèmes que l'on simule, de sorte que la probabilité théorique des événements futurs réside dans la suite de nombres pseudo-aléatoires et non dans la simulation elle-même.

Nous sommes maintenant en mesure d'affronter notre première question : peut-on construire une machine perceptrice ? Oui, bien que celles qui ont été construites jusqu'ici soient assez primitives. J'ai déjà exposé un certain nombre de données sur les performances de l'une des premières d'entre elles au chapitre 9. Cet objet, appelé Darwin III, possède un œil mobile et un bras à quatre articulations équipé de senseurs tactiles sur la partie distale de la dernière articulation, ainsi que des neurones kinesthésiques aux articulations.

Il comporte des neurones simulés dans de nombreux répertoires, présentant une diversité à la fois au niveau de leurs connexions locales et de la force de leurs liaisons synaptiques. Bien qu'il ne se déplace pas, son œil et son bras peuvent effectuer toutes les combinaisons de mouvements possibles, dans les limites imposées par sa structure mécanique. Des objets appartenant à un monde de formes choisies au hasard passent au hasard dans son champ visuel – et, à l'occasion, à la portée de son bras et de son toucher. Les efficacités synaptiques de ses neurones sont définies au départ par un générateur de nombres aléatoires. À la suite de quelques rencontres avec des objets (devant lesquels il réagit), Darwin III présente un comportement qui ressemble beaucoup à une catégorisation perceptive (voir la figure 9-6). Et cela se produira tant que ses circuits neuronaux auront été construits pour répondre à des valeurs (par exemple, au fait que la lumière est préférable à l'obscurité, ou que le contact tactile vaut mieux que l'absence de contact). Ainsi, Darwin III effectue des catégorisations fondées sur des valeurs.

Je me dois de donner ici quelques précisions. Oui, ce qui donne lieu aux valeurs *est* programmé dans la conception de la machine. Mais ces valeurs ne correspondent à *aucune* catégorie, et les catégories, elles, ne sont pas programmées. La programmation des valeurs est permise parce que les valeurs sont considérées comme étant issues, *au cours de l'évolution*, de la sélection d'un type particulier de préférences en raison des avantages sélectifs que celles-ci conféraient aux individus d'une espèce donnée. Si nous voulions simuler un chat, nous y introduirions peut-être des systèmes de valeurs ou des contraintes sur les circuits afin de faire en sorte, par exemple, que les mouvements conduisant le chat à se lécher le poil (action détectée par des parties simples de son propre système nerveux) soient plus susceptibles d'être récompensés. Mais nous finirions quand même par observer, et non pas par programmer, les effets des contraintes sur le comportement. Chez nos automates, nous *n'avons pas* programmé les types de catégorisations qui résultent effectivement de la sélection somatique, parce que ces modifications sont épigénétiques.

À ce propos, les valeurs n'ont rien de particulièrement mystérieux. Par exemple, afin de fournir à Darwin III des valeurs qui soient positives pour les objets éclairés se trouvant dans la partie centrale de son champ de vision, on a construit des neurones spécialisés ayant des connexions d'entrée plus denses avec la partie centrale de sa « rétine » et moins denses avec la partie périphérique.

Ainsi, en tombant sur la partie centrale de la rétine, l'image d'un objet éclairé suscite des réponses marquées de la part de ces neurones. Ces réponses sont ensuite transmises vers le voisinage des synapses reliant les neurones visuels aux neurones moteurs qui commandent les mouvements de l'œil. Dans ces systèmes, l'activité laisse une « trace chimique ». Et, tant que ces traces persisteront, tout mouvement amenant le stimulus vers la région centrale de l'œil sera suivi d'un signal issu des valeurs, entraînant un renforcement probabiliste des liaisons synaptiques concernées. Cela fera donc croître la probabilité que des mouvements semblables se produisent la prochaine fois qu'un stimulus apparaîtra dans une position semblable.

Le comportement de Darwin III est très limité : l'ensemble de caractéristiques des stimulus sur lesquels il effectue des catégorisations n'est pas très étendu. De plus, il n'est pas vraiment capable d'apprendre, bien que certaines expériences, au cours desquelles on l'avait équipé d'un système « gustatif », incitent à penser qu'il est capable de modifier ses modes de sélection d'objets lorsqu'on modifie ses valeurs. Quoi qu'il en soit, pour tester de tels comportements, il nous faut un environnement beaucoup plus riche – un environnement comportant des objets réels. La construction d'une machine perceptrice dont le système nerveux sera simulé à l'aide d'un supercalculateur, mais dont les « vrais » yeux et l'appareil moteur se trouveront dans une « créature » qui « vivra » dans une autre pièce, est actuellement en cours (figure 19-2). Cet objet, appelé NOMAD (*Neurally Organized Multiply Adaptive Device* ; en français, « Engin à structure neuronale et à adaptabilité multiple »), est relié à son cerveau par télévision et par radio. Ce cerveau – plus complexe que celui de Darwin III – est toujours conçu comme un système sélectif. Les tâches que NOMAD devra réaliser associent la catégorisation et l'apprentissage ; par exemple, il aura à choisir, parmi des objets de formes diverses, ceux qu'il considérera comme ayant de la valeur. Contrairement à ce qui se passe avec les robots habituels, ces choix ne seront pas programmés : ils résulteront d'un entraînement.

J'ai donné à l'étude de ce type d'engins le nom de *noétique*, du grec *noein*, qui signifie percevoir. Contrairement aux engins cybernétiques, qui sont capables de s'adapter dans les limites de certaines contraintes fixes de conception, et contrairement aux robots, qui utilisent des principes cybernétiques sous contrôle programmé, les engins noétiques agissent sur leur environnement par des moyens

sélectifs et à travers des catégorisations fondées sur des valeurs. Ce domaine n'en est qu'à ses premiers stades, mais il promet déjà de nous apprendre énormément de choses sur la démarche à adopter pour étudier les niveaux et les boucles des systèmes neuronaux structurés. À terme, la noétique est également appelée à acquérir une importance pratique considérable.

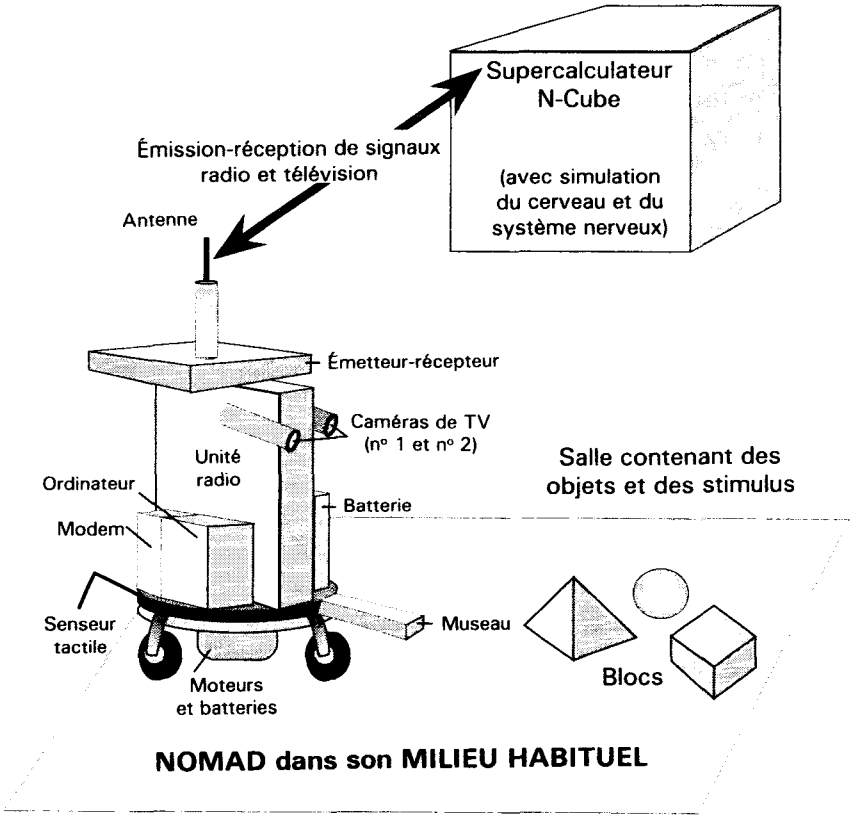


FIGURE 19-2

NOMAD (engin à structure neuronale et à adaptabilité multiple – Neurally Organized Multiply Adaptive Device) est un objet réel construit selon des principes semblables à ceux de Darwin III. Bien que le cerveau de NOMAD soit simulé sur un puissant supercalculateur, ce « cerveau » ne fonctionne pas comme un ordinateur. NOMAD « vit » à l'Institut de Neurosciences, et il est le premier objet non vivant capable d'« apprendre » au sens biologique du terme. Avec son « museau », il ramasse des blocs magnétiques de différentes formes et différentes couleurs qui lui fournissent des « valeurs » (des stimulus électriques) par contact. NOMAD a l'air d'un robot, mais, contrairement aux robots, il n'est pas sous le strict contrôle d'un programme. Il opère comme un engin noétique, un engin à structure neuronale qui fonctionne selon des principes sélectionnistes. Les impulsions nerveuses transmises à NOMAD par son cerveau simulé (impulsions qui, chez un animal, serviraient à activer des muscles) sont traduites en signaux destinés à ses roues par un ordinateur de bord.

Pouvons-nous étendre ces notions à la construction d'un objet doué de conscience primaire ? La réponse n'est pas simple. Mais on peut se risquer à dire qu'elle est très vraisemblablement affirmative. En principe, il n'y a aucune raison pour qu'il soit impossible de simuler, grâce à des principes sélectifs, un cerveau possédant une conscience de ce type – pourvu que la simulation comporte les parties appropriées. Mais il reste encore beaucoup à faire avant que l'on parvienne à concevoir effectivement un objet conscient. Par exemple, personne n'a jusqu'ici réussi à simuler un système cérébral capable de former des concepts, et donc de *reconstruire* des portions de cartographies globales. Cet objectif constitue en lui-même un énorme défi. Et si l'on songe qu'on a besoin par ailleurs de multiples modalités sensorielles, d'appendices moteurs élaborés et d'un grand nombre de neurones simulés, il n'est pas du tout sûr que les supercalculateurs actuellement disponibles et leurs mémoires soient à la hauteur de la tâche.

Si notre modèle de la conscience primaire est correct, une telle simulation, qui en principe est possible, peut nous aider à tester la cohérence interne de ces idées. Comment ? En nous permettant de déterminer si le comportement d'un objet capable de corréler une scène par réentrée entre la mémoire des valeurs-catégories et les catégorisations perceptives l'amène ou non à choisir des combinaisons d'événements extérieurs non reliés de façon causale *de manière à satisfaire ses propres besoins adaptatifs* – c'est-à-dire selon le relief qu'il attribue lui-même aux choses et sa propre histoire. Le test d'efficacité consiste à laisser intacts les circuits de l'objet puis à couper les boucles réentrantes clés, une à une, pour voir quels sont les effets néfastes – s'il y en a – de ces coupures sur le comportement adaptatif de l'objet. (La procédure ressemble un peu au test des dissociations implicite/explicite évoquées au chapitre précédent.) Cet exemple illustre l'un des principaux avantages des simulations, dont l'intérêt a d'ailleurs déjà été démontré dans le cas de machines perceptrices comme Darwin III. Étant donné la complexité des structures neuronales et des comportements, ainsi que leurs multiples niveaux d'interaction, seul un ordinateur extrêmement rapide et doté d'une énorme capacité de mémoire pourrait contenir toutes les configurations dont il est nécessaire de disposer à chaque niveau pour les besoins de l'analyse scientifique. Les ordinateurs ne sont pas des modèles adéquats du cerveau. Néanmoins, ils constituent l'outil heuristique le plus puissant dont nous disposions pour tenter de comprendre la matière de l'esprit.

Suite à ce que je viens de dire, vous avez probablement déjà deviné la réponse à la dernière de nos questions, à savoir celle qui concerne les objets artificiels dotés d'une conscience d'ordre supérieur. Il sera peut-être possible de construire un jour de tels objets, mais actuellement cela est si peu vraisemblable qu'il ne vaut même pas la peine de beaucoup s'y attarder. Non seulement faudra-t-il avoir construit auparavant un objet ayant une conscience primaire, mais nous devrons aussi avoir compris comment *au moins deux* systèmes de ce type pourraient intentionnellement faire en sorte que quelque chose « soit ce qu'elle est pour l'un l'autre sous les auspices d'un symbole », selon les mots de l'écrivain Walker Percy (voir la postface). En d'autres termes, il faudrait que les objets dotés de conscience d'ordre supérieur possèdent un langage et l'équivalent d'un comportement au sein d'une communauté linguistique. Il reste encore beaucoup de choses à comprendre à propos de l'organisation de la mémoire linguistique, et il ne semble pas vraisemblable que l'on parvienne à résoudre rapidement ce problème. Pour l'instant, nous pouvons être tranquilles : nous sommes encore les seuls systèmes connus qui possèdent une conscience d'ordre supérieur fondée sur le langage, et aucun objet ne viendra nous faire concurrence avant longtemps.

En principe, cependant, il n'y a aucune raison de penser que nous ne serons jamais capables de construire de tels objets. Savoir si nous devons le faire ou non est une tout autre histoire. Les questions morales sont truffées de choix difficiles et de conséquences imprévisibles. Et comme nous avons déjà suffisamment à faire concernant la société humaine, je pense que nous pouvons légitimement laisser en suspens, pendant encore quelque temps, les jugements et les réflexions en matière d'objets conscients. Nous avons des tâches bien plus urgentes à accomplir.

En réfléchissant à ce genre de questions, nous ne devons pas perdre de vue à quel point la construction d'une science réellement intégrée de l'esprit en est encore à ses débuts. Bien sûr, la psychologie non expérimentale est l'une des « sciences » les plus anciennes. Mais la psychoneurobiologie est encore dans l'enfance. Par conséquent, nous devrons probablement attendre encore quelque temps – et peut-être même longtemps – avant de voir apparaître le genre de progrès que j'ai évoqués au cours de ce chapitre. Pour ma part, j'ai la conviction qu'on finira un jour par construire effectivement des objets conscients, et que cela se passera dans des conditions d'humanisme éclairé, en faisant dûment attention au bien-être de

l'humanité. Mais cela prendra très longtemps. Le « pop-artist » Andy Warhol m'avait abordé une fois à une fête et m'avait dit qu'il faisait collection de revues scientifiques, mais qu'il était incapable de les comprendre. S'étant éloigné un instant, il revint et me dit : « Je peux vous poser une question ? » « Bien sûr », lui répondis-je. « Pourquoi est-ce que la science est si lente ? », me demanda-t-il. Je lui dis alors : « Monsieur Warhol, quand vous faites un portrait de Marilyn Monroe, est-ce que vous tenez à ce qu'il lui ressemble exactement, ou du moins aussi exactement que possible ? » Il me dit alors : « Oh, non. Et de toute façon, ce sont mes assistants qui le font, à un endroit appelé l'Usine. » Sur quoi je déclarai : « Eh bien, en science, il faut que les choses soient exactes, aussi exactes que possible. » Il me regarda d'un air compatissant et me dit : « Comme c'est affreux. »

L'une des questions que je n'ai pas encore abordées est celle que certains philosophes appellent l'opposition « chauvinisme »/« libéralisme ». *Faut-il* que les objets du genre de ceux décrits ici soient constitués de molécules organiques ? En ce qui concerne les machines perceptrices, nous avons déjà la réponse : c'est non. Mais il *faudra* tout de même imiter de près des structures qui n'existent qu'en biologie. Cependant, si notre vision de l'esprit est correcte, le libéralisme ne sera jamais à l'ordre du jour. Même avec une totale connaissance des structures cérébrales, je parie que nous ne serons jamais capables de mettre au point des logiciels qui permettent à la conscience de fonctionner sur un ordinateur, aussi puissant soit-il, dans les conditions requises par le fonctionnalisme (voir la postface). Les contraintes imposées par la morphologie et la sélection vont à l'encontre de ce genre d'espoirs.

Par conséquent, la réponse à la question posée par le titre de ce chapitre est la suivante : en principe oui, mais les problèmes pratiques posés par la « fabrication » de consciences d'ordre supérieur sont si difficiles à résoudre que nous n'avons pas à nous en préoccuper pour l'instant. Quant à l'idée d'un objet ayant une conscience primaire, la réponse est un « oui » un peu plus ferme, à ceci près que nous avons encore beaucoup à apprendre sur la façon dont les systèmes neuronaux donnent lieu à des concepts dans un corps. Enfin, en ce qui concerne les objets dénués de conscience mais capables d'effectuer des catégorisations, nous avons vu qu'il existe déjà des prototypes de « machines » perceptrices.

En moins de cinquante ans, nous avons fait énormément de chemin en matière d'ordinateurs, et ce en n'imitant qu'*une seule*

fonction cérébrale : la logique. Il n'y a aucune raison de penser que nous échouerons dans nos tentatives d'imiter d'autres fonctions cérébrales au cours des quelque dix ans à venir. Et étant donné les travaux de recherches très prometteurs sur la modélisation du système nerveux à l'aide de systèmes neuronaux synthétiques (du type de ceux utilisés dans Darwin III), nous serons peut-être bientôt en mesure de nous intéresser aux résultats susceptibles d'être obtenus en branchant dix machines perceptrices (P), capables de classer les nouveautés par catégories, à une machine de Turing (T), capable d'effectuer des opérations logiques. Il se peut que la combinaison résultante – ou machine PT – se comporte vis-à-vis de la reconnaissance de la nouveauté, en gros, comme un chasseur accompagné de ses chiens, pourvu que les machines P aient été correctement entraînées et que la machine T ait été correctement programmée par un opérateur humain. Si cette entreprise réussit, les résultats obtenus avec des ordinateurs branchés sur des NOMAD, ou engins noétiques, auront des implications pratiques et sociales profondes. Je ne sais pas combien de temps il nous faudra encore pour parvenir à fabriquer ce genre de choses, mais ce que je sais – comme c'est le cas habituellement en science –, c'est que l'avenir nous réserve certainement de nombreuses surprises.

Chapitre 20

Symétrie et mémoire : aux sources de l'esprit

Ce que l'univers a de plus incompréhensible c'est d'être compréhensible.

Albert Einstein

« L'homme pense ; donc je suis », dit l'Univers.

Paul Valéry

La cosmologie a souvent fait partie des mythes et des sciences de bien des civilisations. L'esprit y a toujours joué un rôle central – qu'il soit intérieur, extérieur ou ultérieur. Il est naturel que des créatures telles que nous se demandent comment tout cela est arrivé, comment il se fait que nous soyons là nous-mêmes, et comment il se fait que nous ayons fini par devenir conscients du monde dans lequel nous nous trouvons.

Dans certaines cultures, les cosmologies religieuses du passé ont été remplacées par une cosmologie scientifique nettement liée aux aspects les plus inaccessibles de la physique théorique. Mais aussi magnifique et aussi mystérieuse soit-elle, cette cosmologie scientifique ne contient aucun principe intrinsèque capable de nous conduire jusqu'à nous-mêmes, en tant qu'observateurs conscients qui formulons la physique et la relions à la cosmologie, et qui ne résistons pas à l'envie de nous intégrer dans la vision scientifique du monde que nous avons construite. Or, même une « théorie de tout », comme l'appellent certains physiciens, serait incomplète si elle ne fournissait pas un tel principe.

Au cours de ce livre, j'ai affirmé que l'esprit était issu de façon très précise des mécanismes d'évolution de la morphologie. J'ai essayé de montrer que la conscience est apparue, du moins dans cet infime fragment de cosmos, à un moment donné de l'histoire. Cependant, le fait qu'elle soit apparue au sein de structures matérielles précises présentes dans le cerveau ne signifie pas que l'on puisse l'identifier à ces structures. En effet, comme nous l'avons vu, la conscience dépend aussi des relations avec l'environnement

et, à son niveau le plus élevé, des symboles et du langage employés par une société donnée.

La conscience d'ordre supérieur engendre un riche domaine cognitif, affectif et créatif – les sentiments (sensations), les pensées, les émotions, la conscience de soi, la volonté, l'imagination. Elle est capable de construire des objets mentaux artificiels tels les fantasmes. Au niveau culturel, elle mène à l'étude des relations stables existant entre les phénomènes (la science) et les objets mentaux (les mathématiques), ainsi qu'à l'étude des relations existant entre les phrases qui font référence à des objets mentaux et à des phénomènes (la logique).

Ma description de la façon dont l'esprit est apparu dans la nature peut paraître étrange. Cela est dû en partie au fait qu'il ne semble pas s'être formé de la même manière que nos constructions et nos inventions les plus chères – c'est-à-dire à travers des relations bien ordonnées, comme celles qui régissent la logique, l'arithmétique et la physique, et qui nous ont permis de construire des ordinateurs et autres systèmes d'information.

Cependant, cela ne veut pas dire qu'il soit impossible de trouver dans la nature un principe profond sous-jacent à l'évolution de la conscience. Au cours de ce chapitre final, je compte me demander quel pourrait bien être ce principe, pour ensuite tenter de le relier à un principe plus fermement établi, dont la plupart des physiciens seraient d'accord pour dire qu'il s'agit de l'un des principes les plus fondamentaux de la physique et de la cosmologie. Enfin, j'essayerai de voir comment ces deux principes pourraient, ensemble, régir la pensée scientifique future ainsi que notre vision de la place que nous occupons dans le cosmos.

Il est certain qu'au cours du prochain siècle – et peut-être même avant – la physique et la biologie seront appelées à « correspondre » l'une avec l'autre d'une façon très intime. Au pire, elles échangeront des vues sur la façon dont l'observateur humain influe sur les mesures physiques, et sur les rapports existant entre les perceptions des observateurs et leurs descriptions physiques. Il s'agit là d'un problème clé en physique quantique (voir la postface).

Ces prédictions peuvent sembler vagues et utopiques, mais je pense que les lecteurs ne devraient se faire une opinion personnelle qu'après avoir réfléchi à ce qui a déjà été dit et à ce qui va être dit ici. Comme c'est le cas pour tous les spécialistes scientifiques, mes connaissances et mon expérience sont parfois insuffisantes pour apprécier les relations existant entre les divers domaines de la

connaissance. On peut même dire que lorsque des chercheurs comparent leurs expertises respectives, cela mène parfois à des impasses. À ce propos, George Uhlenbeck, un grand physicien – l'homme qui m'enseigna la mécanique quantique et la mécanique statistique –, me demanda un jour de lui présenter Fritz Lipmann, un biochimiste tout aussi éminent qui m'avait orienté pendant mon doctorat. J'organisai donc un dîner pour nous trois. Pendant que Fritz mangeait sa soupe avec plaisir, George, qui était un homme très sérieux et digne, dit : « D'après le calcul du potentiel de Gibbs dans diverses phases, j'en déduis que la vie n'est apparue qu'une seule fois au cours de l'histoire de l'univers. » Tout en continuant à avaler méthodiquement sa soupe, Fritz leva la tête entre deux cuillerées et répondit énergiquement, avec son charmant accent : « Za arrife tout le temps. » George lui assena alors des arguments physiques très élaborés. Fritz l'écouta – sans cesser de manger – et finit sa soupe. Reposant la cuiller, il dit : « Quand nous afons broyé le muscle pectoral du picheon, ils ont dit qu'on n'aurait pas la vhozvhorilazion ogzidatife. Quand nous afons eu le microsome, ils ont dit qu'on n'arrifferait pas à faire la synthèse des protéines. On l'a faite. Ne fous en faites pas, za arrife tout le temps. » Ils se sont souri, mais je suis certain que, lorsqu'ils se sont levés de table, chacun d'eux se sentait frustré à cause des limitations intellectuelles inhérentes à la spécialité de l'autre.

Notre siècle a été le témoin de ce qui constitue peut-être la plus grande révolution de tous les temps en matière de pensée scientifique. Et cette révolution ne réside pas simplement dans l'adoption de concepts inusités pour parvenir à comprendre l'univers et ses particules élémentaires. Elle réside aussi dans notre conception de l'observateur scientifique et dans l'extension de la généralité de la pensée scientifique. Depuis la découverte des quanta par Planck, suivie des premières formulations de la théorie quantique, par Bohr et Heisenberg, et de la théorie de la relativité par Einstein, les observateurs ne sont plus considérés comme étant complètement détachés de leurs mesures. En effet, lorsqu'on mesure des grandeurs quantiques, la façon dont l'observateur choisit de disposer ses instruments détermine l'issue de la mesure. Et, en théorie de la relativité, les mesures du temps et de l'espace réalisées par différents observateurs dépendent de la vitesse et de l'accélération relatives de ces derniers. Ainsi, les choix conscients des observateurs dans le premier cas, et leur position physique dans le second, doivent être explicitement pris en compte.

Le résultat des efforts de rationalisation de ces découvertes est bien connu : la mécanique quantique et la théorie de la relativité générale sont les deux plus superbes constructions théoriques du domaine scientifique. Leurs descriptions vont des particules élémentaires les plus minuscules et les plus éphémères aux confins de l'univers mesurable.

Ce qui n'est peut-être pas bien compris à l'extérieur de la communauté des physiciens, c'est le fait qu'un même principe mathématique clé sous-tend ces deux descriptions : le principe de symétrie. Je ne tiens pas à entrer ici dans des détails mathématiques, mais je vais tenter d'en dire suffisamment pour vous donner tout de même un aperçu de ce principe. La symétrie est un exemple spectaculaire de la façon dont un argument mathématique issu d'un raisonnement logique peut être appliqué à des descriptions de la nature et conduire à des idées extrêmement générales. Je voudrais m'attarder quelque peu sur la symétrie parce que j'ai l'intention de la comparer à un autre principe qui selon moi sous-tend l'esprit, et de fait toute la biologie : le principe de mémoire. Plus tard, je tenterai de montrer que le fait de comprendre ces deux principes, qui interagissent d'une façon à la fois harmonieuse et tendue, nous permettra de déterminer plus clairement quelle est la place qu'occupent nos esprits dans la nature.

Nous nous sommes tous familiarisés avec la symétrie au cours de notre vie quotidienne ; en gros, nous sommes des créatures à symétrie bilatérale. Nous savons que notre image dans un miroir possède certaines propriétés, et que notre main droite et notre main gauche sont l'image l'une de l'autre dans un miroir (figure 20-1). Aucune opération chirurgicale au monde ne pourra transformer une main droite en une main gauche sans dégâts, mais pour convertir un gant droit en un gant gauche, il suffit de le retourner (de sorte que l'intérieur se retrouve à l'extérieur et vice versa). Cela suggère que certaines *opérations* sont nécessaires à la mise en évidence de certains types de symétrie.

Les principes de symétrie et les règles qui président à ces opérations font partie intégrante de la théorie mathématique des groupes, qui joue un rôle essentiel dans la construction des théories physiques de pointe. Cette théorie mathématique fut formulée au début du XIXe siècle par un jeune et génial mathématicien français du nom d'Évariste Galois, qui perdit la vie au cours d'un duel, à cause d'une femme, alors qu'il était âgé d'à peine vingt ans et demi.

Les idées de Galois sur les groupes révélaient l'impossibilité de résoudre les équations polynomiales du cinquième degré dans le cas général (la plupart des non-spécialistes se seront sans doute arrêtés aux équations du second degré, ou quadratiques). Mais il se trouve que le domaine d'application de ses idées est, quant à lui, extrêmement général. Le groupe des réflexions dans un miroir est lié à des déformations non continues, comme nous l'avons déjà vu (figure 20-1). Mais d'autres groupes ont à voir avec des symétries continues – par exemple, avec des translations dans l'espace. (Cette théorie fut énoncée de manière spectaculaire, au cours de la deuxième moitié du XIX[e] siècle, par le mathématicien norvégien Sophus Lie.) En général, les objets présentant les symétries d'ordre le plus élevé – tels le cercle dans l'espace à deux dimensions et la sphère dans l'espace à trois dimensions – sont relativement dépourvus de traits distinctifs. Et, en général (mais pas toujours), l'adjonction de caractéristiques à ce type d'objets entraîne une diminution de leur degré de symétrie.

Nous voici à même d'aborder l'une des contraintes formelles qui rendent les lois de la physique extrêmement puissantes. Il s'agit du lien existant entre les idées de symétrie et ce qu'on appelle les lois de conservation de la physique. L'étude de la physique a révélé qu'en mécanique, ainsi que dans les champs électriques et dans les champs de particules, un certain nombre de grandeurs fondamentales sont conservées. La masse-énergie, la quantité de mouvement et le spin sont chacun régis, en effet, par des lois de conservation qui exigent qu'aucun d'eux ne soit créé ni détruit au sein du contexte global d'une description physique. Les charges électriques suivent également une loi de conservation : le nombre de particules chargées positivement existant dans l'univers est égal au nombre de particules chargées négativement. Il existe des lois analogues pour les particules élémentaires. Ainsi, le nombre de protons et d'autres particules est conservé, de même que le nombre d'électrons et de particules associées.

Les conséquences de l'application du principe de symétrie sont réellement superbes, car les différentes lois limitent les formes d'interaction possibles entre ces particules. En d'autres termes, les règles décrivant les *interactions* entre particules sont soumises aux contraintes des principes de conservation. Il en résulte que, dans certains cas, des particules ne peuvent être créées ou détruites que par paires, tandis que dans d'autres des particules peuvent être créées ou détruites sans être soumises à ces contraintes.

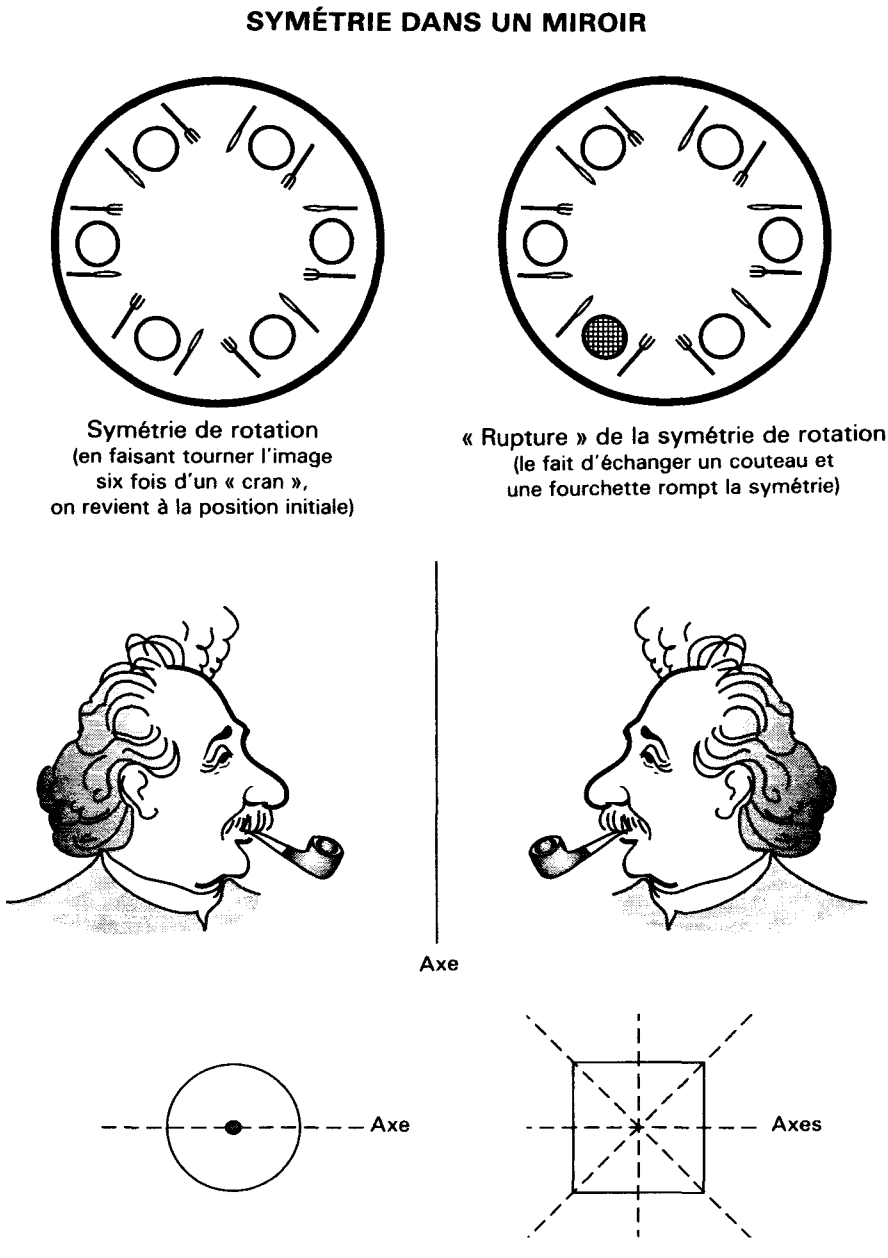


FIGURE 20-1

Divers types de symétries.

Nous arrivons ainsi à l'un des grands thèmes de la physique : *les lois de conservation et la symétrie sont intimement liées.* L'espace vide et le temps sont symétriques ; autrement dit, ils conservent le même aspect malgré toutes sortes de transformations. Ainsi, l'espace demeure inchangé par les translations, les rotations et les changements de direction. Et le temps, lorsqu'il s'inverse, est le même dans les deux sens (j'entends par là le temps des physiciens, et non notre notion individuelle du temps). En fait, en mécanique quantique et en théorie de la relativité, les *lois* du mouvement elles-mêmes sont invariantes par des transformations symétriques telles que la rotation et la translation. Cette invariance garantit que l'issue des événements physiques ne dépend pas des systèmes de coordonnées utilisés pour les mesurer.

Tant qu'on ne leur appliquera aucune force, ni la vitesse ni le sens de déplacement d'un corps ou d'une particule en mouvement (leur quantité de mouvement), pas plus que leur énergie, ne changeront. Le mathématicien allemand Emmy Noether fut le premier à montrer que la conservation de ces grandeurs peut être *formellement* identifiée à des principes de symétrie. Par exemple, la conservation de la quantité de mouvement correspond à la symétrie de l'espace par translation. La conservation du moment angulaire correspond à la symétrie de l'espace par rotation. La conservation de l'énergie correspond à la symétrie du temps par inversion de sens. (L'inversion du temps ne peut pas être effectivement réalisée, mais il est possible de vérifier que les lois physiques sont invariantes par de telles opérations.)

Ce fut Einstein qui, le premier, comprit l'importance de l'invariance des lois de la physique et, donc, l'importance de leur symétrie. En fait, sa théorie de la relativité générale peut être considérée comme un moyen de rechercher les conditions de l'invariance *absolue* !

Plus récemment, une série de découvertes a permis d'entrevoir la possibilité d'unifier toutes les interactions entre particules en une seule théorie – une théorie de grande unification (ou GUT pour *grand unification theory*). Cela n'a pas encore été obtenu pour l'ensemble des quatre forces existant dans la nature – l'interaction faible, l'interaction forte, l'interaction magnétique et l'interaction gravitationnelle. Mais des théories partielles spectaculaires ont été énoncées, des théories qui auraient été inimaginables il y a à peine 20 ans. Et on peut dire que s'il existe un langage majeur dans ces théories, c'est sans aucun doute le langage de la symétrie. Les

physiciens espèrent qu'un jour il sera possible de décrire l'ensemble de la nature (c'est-à-dire de la physique) en fonction d'une symétrie qui permettra de rendre compte de tous les champs et de toutes les forces de la même manière.

Nous ne pourrions aborder toutes ces questions sans trop nous écarter de notre sujet. Cependant, je voudrais évoquer deux concepts essentiels aux efforts déployés par les physiciens pour construire une théorie unifiée des champs. Ces concepts – qui sont également essentiels à la cosmologie moderne – sont ceux de symétrie de jauge locale et de rupture spontanée de symétrie. La symétrie locale peut être définie par opposition à la symétrie globale. Pour qu'une transformation laisse invariante une symétrie globale dans un domaine donné, elle doit se produire *partout*. La symétrie locale, en revanche, permet à différentes transformations de se produire dans différentes parties de l'espace et du temps. La théorie de la symétrie locale mise au point par C. N. Yang et F. E. Mills a été décisive pour le succès des tentatives ultérieures d'unification. Par exemple, étant donné un champ, si l'on veut qu'il soit invariant par un changement de symétrie locale, il faut que les choses se passent de telle sorte qu'un *autre* champ agisse pour compenser exactement toutes les modifications locales introduites par la première opération.

Pour comprendre l'essence de la rupture de symétrie, considérons le fond d'une bouteille de vin vide : il a la forme d'une coupole symétrique entourée d'une rigole destinée à recevoir les dépôts. Si l'on pose une boule exactement sur le sommet de la coupole, la situation reste complètement symétrique. Mais la boule rompra spontanément cette symétrie s'il lui arrive de tomber au fond de la rigole – qui représente le minimum d'énergie. La symétrie globale sera ainsi spontanément abolie, bien que la bouteille et la boule continuent de conserver chacune leur symétrie individuelle. Appliquée à n'importe quelle théorie physique, cette idée suggère qu'une *solution particulière* des équations de la théorie peut être moins symétrique que la théorie elle-même. Ce sont des notions telles que celles-ci qui sous-tendent la théorie électrofaible et la théorie des interactions fortes, deux triomphes théoriques récents de la physique moderne.

Une autre découverte de ce siècle est le fait que les lois de la physique sont étonnamment générales. Ainsi, les concepts de symétrie s'appliquent même à des théories sur la façon dont l'univers tel que nous le connaissons aujourd'hui s'est formé. Et, lorsqu'on examine les théories cosmologiques modernes (par exemple, la théo-

rie de l'inflation de l'univers et celle du Big Bang), on s'aperçoit que l'univers, en tant que fonction de la température (qui diminuait) et du temps (qui augmentait), a donné lieu aux particules élémentaires, au cours de son évolution, à travers des phénomènes de rupture de symétrie. À un moment donné, longtemps après le début de ce processus (et longtemps après la formation des particules et des champs que nous connaissons), les galaxies, les étoiles et notre système solaire avec ses planètes apparurent. Sur Terre, par un processus dont les détails ne sont toujours pas connus, la vie prit forme et l'évolution commença, ce qui finit par aboutir à l'émergence de l'esprit. Fritz Lipmann aurait tant aimé savoir quel était ce processus !

Quel nouveau principe sous-jacent au développement de l'esprit et de l'intentionnalité au cours de l'évolution pouvons-nous proposer dans ce contexte ? Je suggère que ce nouveau principe est un principe de mémoire – un principe qui revêt de multiples formes, mais qui possède des caractéristiques générales qui se retrouvent dans toutes ses variantes. J'utilise ici le mot « mémoire » dans un sens plus général que d'habitude. La mémoire est un processus qui n'est apparu qu'avec la vie et l'évolution, lorsque celles-ci donnèrent lieu aux systèmes décrits par les sciences de la reconnaissance. Tel que je l'utilise ici, le mot mémoire évoque certains aspects de l'hérédité, des réponses immunitaires, de l'apprentissage réflexe, du vrai apprentissage consécutif à la catégorisation perceptive et des diverses formes de conscience (figure 20-2).

Dans tous ces cas, des structures sont apparues qui ont permis d'établir des corrélations significatives entre des configurations dynamiques actuelles, en cours, et celles imposées par des configurations passées. Ces structures sont toutes différentes, et les propriétés de la mémoire dépendent du système dans lequel elle se manifeste. Cependant, ce que tous les systèmes de mémoire ont en commun, c'est d'être évolutifs et sélectifs. La mémoire est une propriété essentielle des systèmes biologiques adaptatifs.

Cette extension du terme peut sembler désespérément vague. Mais voyons ce que tous ces phénomènes ont en commun, car en fait, il s'agit de quelque chose de très spécifique. Ce qu'ils ont en commun, c'est leur *relative stabilité de structure au cours des processus sélectifs de cartographie.* Par souci de clarté, je tiens à dire ici quelques mots sur la structure, la stabilité et les cartographies. La loi physique qui parle de structure et de stabilité est le second principe de la thermodynamique. Cette loi stipule que l'entropie, qui mesure

le désordre présent dans un système, peut spontanément augmenter ou rester constante dans un système isolé, mais que jamais elle ne diminue. (Par système isolé, j'entends un système où il n'entre et d'où il ne sort aucune matière ni énergie.) Le système le plus ordonné possible est celui d'un cristal parfait (un cristal dont les atomes forment un réseau symétrique dont la largeur de maille est toujours exactement la même) au zéro absolu de température.

Depuis les tout débuts de son évolution, l'entropie de l'univers n'a cessé de croître. Mais dans certaines parties de l'univers, qui constituent des systèmes *non isolés* (nous-mêmes, par exemple), l'entropie peut en fait *diminuer* localement en raison des transferts de matière et d'énergie. Diverses interactions chimiques donnent alors naissance à des structures stables, et notamment aux molécules des êtres vivants. La stabilité des structures et leurs transactions énergétiques sont régies par les lois de la thermodynamique, et en particulier par le second principe. Et on sait aujourd'hui que des structures chimiques stables peuvent exister en l'absence de toute forme de vie. En effet, les données d'observation indiquent que des molécules organiques semblables à celles de notre corps – des molécules formées par collision d'atomes d'azote, d'oxygène et de carbone, par exemple – existent même dans l'espace intersidéral. Les conditions de leur formation et de leur dissociation, ainsi que celles de leur stabilité, sont déterminées par l'énergie et l'entropie. Cependant, aussi stables soient-elles, il manque à ces molécules un principe héréditaire. Elles ne font preuve d'aucune capacité de réplication – d'aucune capacité de fabriquer des molécules, que l'on pourrait considérer comme leur progéniture, en utilisant leur propre structure comme matrice.

Je voudrais que les choses soient claires en ce qui concerne ma façon d'utiliser le mot stabilité par rapport à la mémoire. Après tout, il existe des cristaux périodiques en dehors du monde vivant (dans les roches, par exemple), qui grandissent en additionnant des atomes à leur structure selon des règles de symétrie fixes. Mais en fait, il ne s'agit pas là d'une *réplication* ; il s'agit d'une *croissance.* Où est donc la différence ?

La différence est que les systèmes réplicatifs contiennent une structure *apériodique* qui subit une sorte de cartographie ; songez aux séquences d'ADN du chapitre 6. Une réaction chimique recopie fidèlement cette structure apériodique, ce qui aboutit à des structures filles. Mais cette fidélité n'est pas infaillible ; il peut apparaître des structures mutantes, qui seront également recopiées. Il en résulte

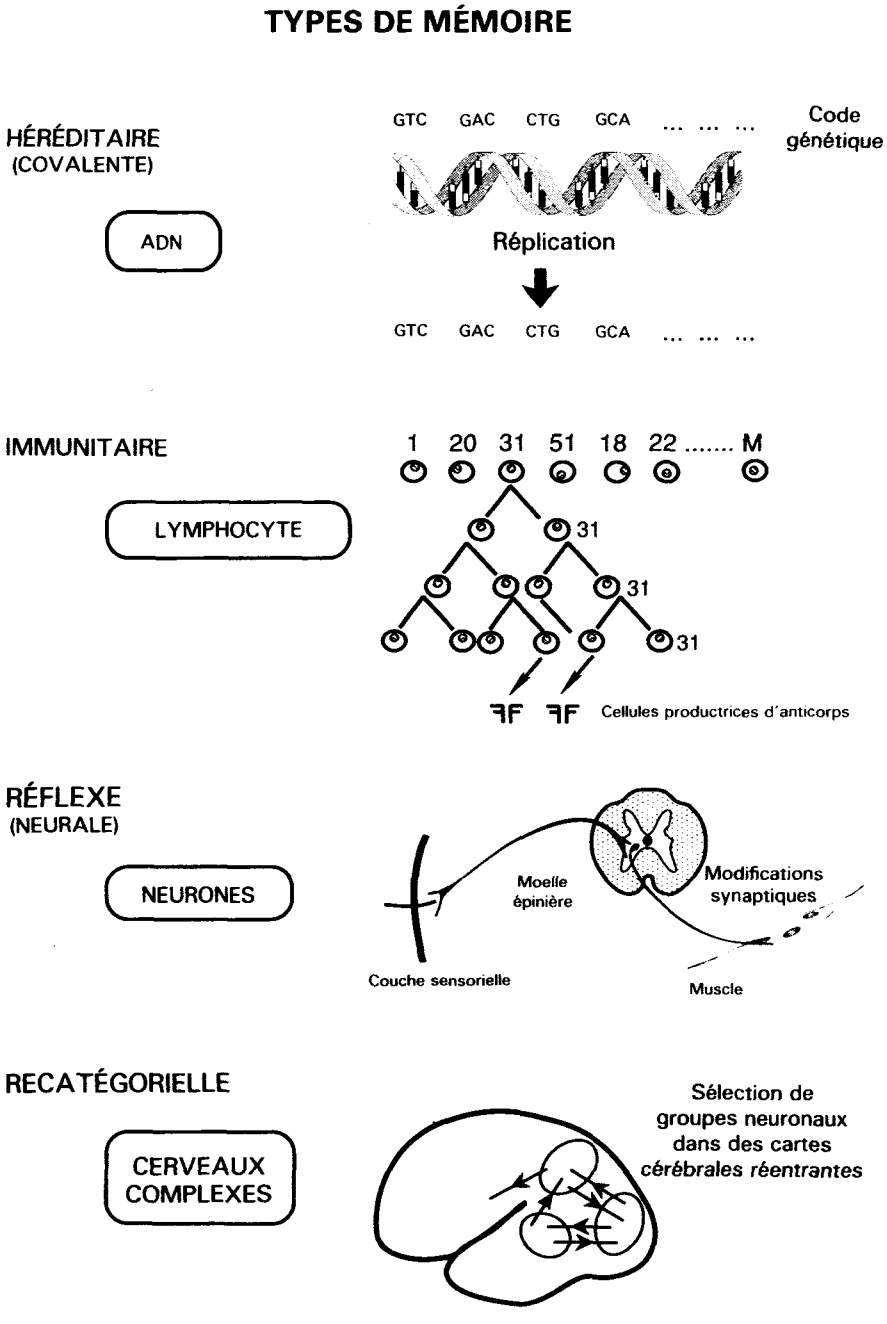


FIGURE 20-2

Divers types de mémoires.

donc une population contenant des variantes. Enfin, la structure apériodique stable est cartographiée, à travers d'autres réactions chimiques, pour servir à fabriquer d'autres types de structures qui la contiennent, de sorte que les variantes favorables auront un avantage sélectif lorsque des copies supplémentaires en seront faites.

Cette description abstraite est celle d'un système vivant : un système auto-réplicatif soumis à la sélection naturelle (voir le chapitre 6). La structure apériodique est soit l'ADN, soit l'ARN, et les structures qui les contiennent ne sont autres que les divers produits protéiques. Cependant, vous remarquerez que le principal processus à l'œuvre ici – qui est absent chez les formes non vivantes – est le principe d'hérédité. Remarquez également que ce principe d'hérédité, qui permet à la proportion de variantes favorisées d'augmenter dans la population au cours du temps, dépend de la stabilité des liaisons chimiques. Dans le cas de l'ADN, il s'agit de celle des liaisons covalentes qui, en reliant entre elles les différentes bases nucléotidiques, donnent lieu au code génétique, dont les « mots » sont des triplets de bases, ou codons. (Chaque codon correspond à l'un des vingt acides aminés qui entrent dans la composition des chaînes protéiques.)

Ainsi, les conditions d'énergie et d'entropie qui règnent dans la fourchette de températures où la vie s'épanouit garantissent le déroulement d'un processus héréditaire. Mais ce sont les phénomènes historiques de *sélection* qui aboutissent aux séquences effectivement présentes dans la population.

L'apparition de ce processus héréditaire constitue un nouveau type d'événement – une forme de mémoire. En effet, hormis l'introduction, dans la séquence, des variations qui se sont révélées favorables, c'est la capacité de conserver en grande partie l'ordre ou topographie de la structure apériodique parente qui permet à ces systèmes de se perpétuer ; autrement dit, ces systèmes conservent une stabilité de structure au cours des phénomènes sélectifs de cartographie. Mais remarquez cependant que cette « mémoire » n'est pas parfaite (alors qu'elle doit l'être dans les messages informatiques). Et en fait, dans une certaine mesure, on peut même dire qu'il *faut* qu'elle contienne des erreurs (des variations d'entropie), ou mutants, pour que le système soit sélectif – pour que, grâce à la variabilité présente dans la population, il soit capable de réagir de façon adaptative à des événements imprévus survenant dans l'environnement.

À mesure que ces structures évoluaient, et que des populations

cellulaires se constituaient en animaux possédant de nombreuses cellules interconnectées et un système nerveux, une nouvelle sorte de mémoire émergea. Cela se produisit à l'issue de modifications synaptiques dans le système nerveux de ces animaux. En raison de la sélection des groupes neuronaux, les comportements qui se révélaient adaptatifs pouvaient être stabilisés par sélection au cours de la vie d'un animal. L'existence d'une mémoire fondée sur les modifications synaptiques est essentielle à de tels comportements.

Chez les vertébrés, le fait que le système immunitaire ait eu besoin de savoir distinguer le soi du non-soi aboutit à la sélection des individus qui possédaient une variante du gène codant pour la molécule d'adhérence des cellules nerveuses ou N-CAM. L'introduction d'une variabilité somatique dans ce qui allait donner lieu aux molécules d'immunoglobuline (anticorps), et le fait que ce processus ait été allié à la réplication fidèle des cellules sélectionnées par des molécules étrangères, fit apparaître un nouveau système de reconnaissance (voir la figure 8-1). Ce système possédait une mémoire immunitaire : la sélection des lymphocytes par des antigènes entraînait des modifications que l'individu conservait ensuite pendant toute sa vie.

De même, ce fut l'élaboration évolutive de récepteurs sensoriels et de couches motrices chez des animaux ayant des cartes cérébrales de plus en plus complexes qui rendit possible l'apparition d'une mémoire fondée sur la catégorisation perceptive. Et, avec l'avènement de capacités conceptuelles et de cartographies encore plus élaborées, les modifications synaptiques survenues au sein des populations de groupes neuronaux en réponse à la nouveauté aboutirent à de nouvelles formes de mémoire.

Chaque type de mémoire reflète une propriété systémique d'un système de sélection somatique. Et chacune de ces propriétés assure une fonction différente, rendue possible par l'apparition, au cours de l'évolution, de la structure neuro-anatomique appropriée. Ces systèmes d'ordre supérieur sont sélectifs et ils se fondent sur les réponses, face aux nouveautés survenant dans l'environnement, de populations de groupes neuronaux structurées en cartes. Ce sont des systèmes de reconnaissance.

À un moment transcendant de l'évolution, une variante qui possédait un circuit réentrant reliant la mémoire des valeurs-catégories aux couples de classification apparut. À ce moment-là, la mémoire devint le substrat et le serviteur de la conscience. Avec l'apparition du langage dans l'espèce *Homo sapiens*, l'itération de

ce même principe dans des mémoires linguistiques spécialisées rendit possible l'avènement de la conscience d'ordre supérieur. Et, au sein de la culture, la conscience d'ordre supérieur finit par donner lieu à une description scientifique de la nature, une description qui nous permet d'étudier les origines de notre propre existence dans l'univers.

Cette description du développement de la mémoire est si différente de la précédente – qui décrivait le développement du cosmos d'après des principes de symétrie – qu'elle peut sembler incommensurable avec elle. L'histoire biologique est une saga *locale*, qui jusqu'ici ne se raconte que sur Terre : elle est historique, elle se déroule dans une fourchette de températures très étroite, elle est extraordinairement complexe et spécifique de certaines structures, elle revêt des formes imprévues et diverses, et elle donne le vertige à ceux qui veulent l'examiner en détail. Néanmoins, elle a débuté dans un monde régi par la symétrie. Et ce n'est qu'à travers la rupture de cette symétrie, avec l'apparition de la chimie et la formation de grosses molécules stables, et à travers des événements sélectifs irréversibles et l'évolution des moyens décrits par les sciences de la reconnaissance, que la mémoire a pu donner lieu à l'esprit. Ce sont les principes de symétrie qui déterminent les possibilités d'apparition de la mémoire, mais ce n'est qu'après la rupture de symétrie, avec l'apparition d'une chimie et d'êtres vivants évolutifs que la mémoire a pu se développer.

La mémoire est à la base de la signification. Et, grâce aux transformations sémantiques rendues possibles par l'incarnation des concepts telle qu'elle est décrite dans la TSGN, il devint possible de développer, au cours de l'histoire culturelle humaine, d'authentiques systèmes de traitement de l'information. Le développement des sciences au cours de l'histoire, par transmission sociale au sein de la culture, nous a permis de reconstituer la vérité à travers des enchaînements de connaissances (voir la figure 14-1). Mais contrairement au développement de la mémoire, cette transmission explosive n'est plus darwinienne. En raison de la nature des systèmes d'information et de celle de la signification elle-même, elle suit des règles lamarckiennes. C'est le fait de s'en *servir* qui entraîne le transfert du contenu des systèmes d'information, ce qui ne fait appel à aucun principe génétique héréditaire. Le transfert s'effectue vers des systèmes somatiques, chacun d'eux étant unique, et les résultats sont spectaculaires – l'esprit humain a opéré sur l'envi-

ronnement des transformations à la fois utiles et horribles. De quoi s'enorgueillir, mais aussi de quoi redoubler de prudence.

J'ai essayé, au cours de ce livre, de mettre au point une vision de l'esprit fondée sur les données scientifiques. Dans l'état actuel de nos connaissances, cette vision ne peut qu'être spéculative. Mais bien qu'elle ait des conséquences philosophiques, elle demeure essentiellement scientifique, et donc sujette à infirmation. Et, malgré le peu que nous sachions sur les détails des mécanismes cérébraux, je pense qu'il est important d'encourager dès à présent la formulation de ce genre de théories. Il y en aura bien une qui nous guidera dans la bonne direction pendant quelque temps. C'est là le maximum que l'on puisse espérer d'une théorie scientifique, hormis le pouvoir de compréhension qu'elle confère. Une théorie n'existe que pour servir de base à une autre, encore meilleure.

Même à la dernière minute, il est bon de rappeler encore une fois que ce que j'ai exposé ici *est* une théorie – et qu'elle n'est pas encore établie. Il reste encore à la tester rigoureusement, et j'ai suggéré des façons de le faire dans ma précédente trilogie sur la morphologie et l'esprit. Comme tous les théoriciens que j'ai pu connaître, je continuerai à penser que ma théorie est correcte jusqu'à preuve du contraire. En matière de formulation de réussites théoriques, l'unité de sélection est habituellement le chercheur mort. Pour cent d'entre nous qui meurent convaincus que ce qu'ils avancent est vrai, un seul a effectivement raison. Et le succès vient encore plus rarement à un chercheur de son vivant. Mais, malgré le risque que cela comporte, chacun de nous doit agir comme si les théories étaient aussi importantes que n'importe quelle autre entreprise scientifique. L'espoir et la conviction sont aussi importants en science qu'ailleurs, à ceci près qu'en science, ils doivent s'effacer devant les expériences.

La théorie de l'esprit que j'ai proposée ici rejette l'idée que la connaissance puisse être absolument certaine. Mais, étant donné l'histoire des succès scientifiques depuis trois siècles, cela ne doit pas nous décourager. Si le cours futur de la science est un tant soit peu déterminé par sa portée actuelle, nous pouvons nous attendre à voir apparaître une remarquable synthèse au cours du siècle prochain. Mais une « théorie de tout » devra certainement inclure à la fois une théorie de l'esprit et une théorie plus complète de l'observateur. La physique et les neurosciences devront donc s'unir pour rendre compte plus pleinement de la relation qui existe entre

les principes de symétrie et de mémoire. Leur coexistence sera à la fois harmonieuse et tendue, mais c'est cette harmonie qui nous permettra de comprendre le monde, les observateurs humains et la place de ceux-ci dans le monde.

Épilogue

J'ai commencé ce livre en vous disant qu'il traitait du sujet le plus important que l'on puisse imaginer. Cette affirmation est évidemment vraie en ce sens que, sans esprit, il n'y aurait ni sujets (vous ou moi) ni sujets de réflexion. Mais j'espère que notre voyage à travers les niveaux et les boucles – des molécules à l'esprit, et en sens inverse de l'esprit jusqu'aux particules élémentaires elles-mêmes – vous aura convaincu d'un aspect moins évident de l'importance de la neurobiologie : le fait que, si nous ne parvenons pas à comprendre comment l'esprit se fonde sur la matière, il subsistera toujours un immense abîme entre nos connaissances scientifiques et notre connaissance de nous-mêmes.

Il n'est pas impossible de jeter un pont au-dessus de cet abîme. Mais la biologie et la psychologie nous enseignent que ce pont comporte de nombreuses parties. La réponse à la question de savoir comment nous faisons pour savoir, pour ressentir, et aussi ce que cela veut dire d'être conscient, ne tient pas dans un unique énoncé philosophique, aussi profond soit-il. Elle n'apparaîtra que lorsque nous aurons compris comment les systèmes et les relations biologiques sont apparus dans le monde physique.

Lorsque l'évolution a donné lieu au langage, le monde imaginable est devenu infini. Le fait de prendre conscience du caractère illimité de l'imagination revêt une très grande beauté et porte en lui beaucoup d'espoirs. Mais si nous voulons comprendre quelle est notre place, en tant qu'observateurs conscients, au sein de nos propres

descriptions, nous devons continuellement revenir de ce monde imaginaire vers le monde matériel. L'un des plus grands défis scientifiques du futur consistera précisément à analyser cette place.

Il serait absurde de vouloir prédire la forme que revêtira une telle science. Mais il est à la fois suffisant et réconfortant de savoir que, quelle que soit cette forme, la vie consciente qu'elle décrira sera toujours plus riche que la description elle-même.

L'esprit sans la biologie : postface critique

Personne n'aime passer trop de temps à émettre des critiques alors qu'il y a du travail créatif à faire. Mais afin d'expliquer pourquoi il est nécessaire d'avoir une théorie biologique du type de celle qui a été formulée dans ce livre, je suis obligé d'asséner quelques coups à plusieurs idées reçues et points de vue établis. Comme je l'ai déjà dit dans le corps de ce livre, un certain nombre de conceptions dominantes sur la conscience et sur l'esprit, même parmi les mieux établies, sont tout simplement intenables. Mais alors, pourquoi s'en préoccuper ? Pour deux raisons. Premièrement, ces conceptions sont dangereusement séduisantes et, tôt ou tard, même le lecteur non initié finira par tomber sur l'une ou l'autre de leurs variantes. Et deuxièmement, l'analyse critique de ces notions nous aidera à mieux définir la nature de notre tâche, qui consiste à montrer comment s'incarne l'esprit.

Il y a encore une troisième raison : aussi erronés soient-ils, quels que soient leurs défauts, ces points de vue – l'idée que la clé se trouve peut-être dans des théories physiques étranges, que le cerveau est un ordinateur, que nous naissons avec une sorte de machine à parler dans la tête – sont intéressants. Mais pour vous convaincre de leur intérêt, je dois d'abord exposer un certain nombre de détails ennuyeux et d'arguments plutôt abstraits qui, si je les avais insérés dans le corps du livre, auraient interrompu le cours de mes descriptions sur la biologie du cerveau. C'est pour cela que j'ai décidé de rassembler les critiques que j'oppose à ces notions dans cette postface.

Mon but est de dissiper l'idée selon laquelle l'esprit peut être compris sans la biologie. Il ne s'agit pas de présenter ici des pensées après coup, mais des extensions des arguments que j'ai exposés dans le corps du livre, destinées non seulement aux spécialistes, mais aussi à tous les lecteurs qui ont envie d'en savoir plus.

Les lecteurs ne devront pas être surpris de constater que cette discussion englobe de très nombreuses disciplines et saute de l'une à l'autre. Les plus difficiles à saisir sont peut-être les sciences cognitives et la linguistique, car elles sont toutes deux abstraites et multidisciplinaires. Mais une fois les obstacles surmontés, on découvre qu'elles sont également fascinantes et extrêmement stimulantes. Cependant, avant de les aborder, tournons-nous une fois de plus vers la physique.

LA PHYSIQUE : L'ÉPOUVANTAIL DE SERVICE

Un épouvantail, c'est entre autres un spectre, une apparition qui nous hante ou nous terrorise. Il peut paraître étrange de dire de la physique, la plus rationnelle des sciences, qu'il s'agit d'un spectre. Mais c'est effectivement ce que devient la physique lorsqu'on l'applique directement à l'esprit. Voici ce que je veux dire.

L'une des façons de résoudre le dilemme posé par l'incarnation de l'esprit et par les apparents mystères de la conscience consiste à considérer que l'esprit et la conscience sont des propriétés directes de la matière. Sous sa forme la plus extrême, cela donne une doctrine philosophique appelée panpsychisme. Selon le panpsychisme, toute matière, même la plus infime particule, est un peu consciente – et cela vaut aussi pour l'univers dans son ensemble. Après tout, raisonne-t-on, nous souhaitons pouvoir dire que l'esprit et la matière sont liés. Et, si nous parvenons à rassembler comme il faut un nombre suffisant de particules très légèrement conscientes, le résultat final sera un être humain conscient. Mais ce point de vue ne précise pas comment déterminer qu'une particule est consciente, sans parler des êtres humains.

Cette position « scientise » un autre point de vue, initialement fondé sur la philosophie idéaliste. Selon ce point de vue, le monde n'est perçu qu'à travers l'esprit, et il se peut donc que la matière n'existe pas – que seul l'esprit existe, comme l'a suggéré Berkeley. Ayant entendu cela, le Dr Johnson donna un coup de pied dans

une pierre et déclara : « Je le réfute ainsi [1]. » Mais la théorie de l'évolution nous fournit une réfutation encore meilleure : puisque la sélection naturelle donne naissance à des êtres sensibles, on voit mal comment l'environnement sélecteur *et* le cerveau pourraient tous deux être des événements mentaux se déroulant dans la tête d'un unique animal sensible dont la progéniture serait également soumise à la sélection naturelle. L'esprit vacille en essayant de comprendre comment les choses auraient pu devenir si compliquées.

La théorie de la sélection naturelle a été tout aussi néfaste au concept idéaliste d'essentialisme chez Platon – l'idée selon laquelle il existe un monde d'essences parfaites, dont les exemplaires dans le monde réel ne sont que le reflet imparfait. Les espèces ne sont ni des essences ni des types ; elles résultent d'une sélection opérant sur des variantes.

Un certain nombre d'individus très intelligents ont été attirés par le panpsychisme, l'idéalisme et l'essentialisme. L'un d'eux fut le poète irlandais William Butler Yeats, auteur d'un traité mystique intitulé *Une Vision* et de quelques poèmes extraordinaires reflétant ses pensées à propos de l'occulte. Avoir de la cervelle et des talents intellectuels ne protège pas contre l'attraction exercée par le surnaturel et le mystique. Dans certaines circonstances, il peut être réconfortant de croire à ce genre de choses, et en particulier si l'on tient absolument à l'idée d'immortalité. Mais, comme l'a dit ma mère sur son lit de mort : « Je ne suis pas pressée. » Lorsqu'on lui demanda pourquoi, elle sourit et dit : « Parce que personne n'en est jamais revenu me dire qu'on s'y amuse comme des fous. »

La plupart des bons physiciens ne croient guère aux idées du panpsychisme et aux esprits désincarnés. Mais certains très bons physiciens ont cependant recherché les réponses au mystère de la conscience au-delà des faits biologiques. Par exemple, certains ont supposé que ces réponses résideront dans de nouvelles théories physiques, et notamment dans une théorie de la gravitation quantique. Pour expliquer pourquoi ils ont été tentés de le faire, et pourquoi je pense que cela revient à faire appel à un épouvantail de service, il me faut dire quelques mots de plus sur les différences entre la physique et la biologie.

La physique est la mère de toutes les sciences : la plus ancienne,

1. D'après un passage de la *Vie de Samuel Johnson*, dont l'auteur est le mémorialiste britannique James Boswell, et qui retrace fidèlement, jour après jour, les propos tenus pendant vingt ans par le « Dr Johnson », moraliste, critique et lexicographe anglais.

la plus fondamentale, celle dont la portée est la plus générale. Elle diffère de la biologie par sa généralité : elle s'applique aussi bien à tous les objets ayant une intentionnalité (y compris les êtres humains) qu'à ceux qui en sont dépourvus. En revanche, la biologie telle que nous la connaissons est une science spécifique. Elle s'intéresse à des phénomènes qui se déroulent à l'intérieur d'une fourchette très étroite de températures (ou d'énergie) et de pression, et qui dépendent d'une chimie très particulière. Encore plus spécifique est le fait que la biologie soit historique. L'évolution est fondée sur une séquence historique *particulière* de sélections naturelles survenant au sein de populations d'organismes diversifiés. En revanche, pour formuler les lois générales de la physique, il n'est nul besoin de considérer ce genre de choses.

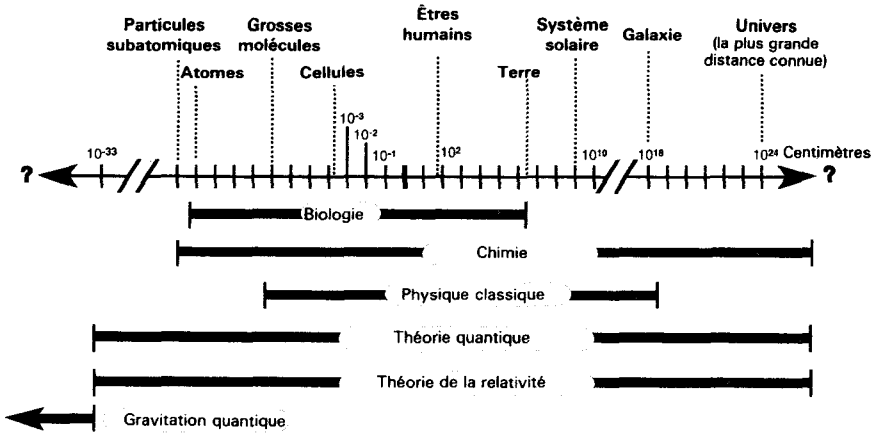
Ce siècle a été le témoin d'une étonnante révolution intellectuelle fondée d'une part, sur la découverte de Planck selon laquelle la matière rayonne de l'énergie par paquets, ou quanta, finis et discrets, et d'autre part, sur la théorie de la relativité d'Einstein, qui remplaça le temps et l'espace par la notion d'espace-temps et introduisit l'idée que la gravitation et la matière représentent la courbure d'une variété spatio-temporelle de dimension quatre. L'élaboration de ces idées révolutionnaires transforma nos idées sur la mesure (figure P-1) et remit profondément en cause les notions habituelles de simultanéité des événements et de causalité.

Ces notions habituelles furent remplacées par des notions étranges, ou du moins peu familières. Les travaux de Planck et d'Einstein débouchèrent également sur un certain nombre de problèmes extraordinaires qui, à ce jour, n'ont pas encore été résolus. Leur « étrangeté » a amené un certain nombre de chercheurs à suggérer que le problème de la conscience y était mêlé.

Les idées derrière ces lois physiques de base sont parfois vraiment étranges (lire « non familières », dans le sens de « contraires à l'intuition »). Contrairement aux idées de la biologie, elles sont *très* générales et, souvent, la meilleure façon de les exprimer consiste à faire appel à des théories mathématiques très puissantes et d'une grande beauté. La symétrie, que j'ai abordée au chapitre 20, en est un exemple.

Étant donné leur généralité et leur pouvoir de prédiction, ces notions de physique sont très séduisantes. Cependant, malgré leur puissance, de profonds problèmes surgissent lorsqu'il s'agit d'en comprendre l'application. Un exemple de cela nous vient de la théorie de la mesure quantique, qui doit être prise en compte

lorsqu'on essaye de mesurer la position ou la quantité de mouvement d'une particule élémentaire. Confrontés aux paradoxes qui apparaissent lors de ces tentatives, d'éminents mathématiciens tels que John von Neumann, et de non moins éminents physiciens tels qu'Eugene Wigner, furent tentés de suggérer que la conscience elle-même intervenait de façon causale dans le processus de mesure quantique.



Échelles de grandeur naturelles et théories applicables

FIGURE P-1

Les échelles de grandeurs naturelles définies par la physique. Au-dessous de 10^{-33} cm, la théorie de la relativité et la théorie quantique ne s'appliquent plus. Au niveau des molécules et au-dessous, la théorie quantique est essentielle (et bien entendu, elle s'applique à toutes les échelles au-dessus de 10^{-33} cm). Dans le domaine des grandes vitesses et des grandes accélérations, on est obligé d'appliquer la théorie de la relativité. Mais au niveau des objets macroscopiques (et en particulier en ce qui concerne les humains et leur cerveau), les descriptions classiques fournissent une très bonne approximation de la réalité ; la convergence des théories quantique et classique aux grandes échelles est connue sous le nom de principe de correspondance. Remarquez que la gamme de tailles du cerveau et la fourchette de températures dans laquelle on trouve des êtres vivants sont toutes deux très étroites. Les échelles sont représentées ici par des puissances de dix – autrement dit, selon une graduation logarithmique.

Nombreuses sont les questions liées à ces propositions, et en parler ici nous mènerait trop loin. Mais permettez-moi tout de même d'esquisser l'un des aspects du problème de la mesure quantique afin de montrer *pourquoi* ces scientifiques ont été tentés d'introduire la conscience dans la physique. Je vous prie cependant de ne pas oublier, pendant ce qui suit, que la physique s'intéresse aux corrélations formelles existant entre les propriétés les plus générales des choses dans l'espace-temps. Les théories de la physique ne s'intéressent *pas* aux capacités sensorielles proprement dites, ni à

la catégorisation des objets macroscopiques identifiables, ni à l'intentionnalité. Cependant, lorsqu'on plonge dans la mécanique quantique, on oublie facilement ces restrictions, parce que les décisions de l'observateur paraissent affecter les mesures qu'il ou elle effectue. Pour comprendre pourquoi, considérons quelques-unes des principales caractéristiques de la théorie quantique.

La théorie quantique est la théorie dont l'applicabilité est la plus générale. Et, comme elle a affaire à d'énormes quantités d'énergie et à de minuscules particules, elle a mis en évidence des comportements qui vont à l'encontre de ce à quoi on s'attend. Ainsi, les particules ne peuvent pas être distinguées les unes des autres. En outre, elles présentent un comportement dual : dans certaines conditions, la meilleure façon de les décrire consiste à les représenter comme des ondes, et dans d'autres comme des particules. En fait – comme Max Born fut le premier à le suggérer –, ψ la fonction d'onde psi, qui apparaît dans l'équation fondamentale de Schrödinger, mesure, lorsqu'on en prend la valeur absolue et qu'on l'élève au carré, la probabilité de trouver une particule à un endroit donné, n'importe où dans l'espace !

Si cependant on tente de déterminer cette position à l'aide d'une expérience, on perd à tout jamais la possibilité de déterminer la quantité de mouvement de la particule avec la même précision. Ce principe, dit principe d'incertitude d'Heisenberg, est fondamental : il existe une relation conjuguée entre la position et la quantité de mouvement (la masse multipliée par la vitesse) de toute particule, et cette relation dit que la précision du produit de ces variables ne peut pas être inférieure à la constante de Planck. Et cela n'est *pas* simplement dû au fait que, pour mesurer précisément la position d'une particule, on soit obligé d'utiliser des particules ou des ondes de longueur d'onde beaucoup plus petite, ayant donc une énergie beaucoup plus grande – ce qui donne, inévitablement, « un coup de pouce » à la quantité de mouvement de la particule. Il s'agit là d'une propriété *fondamentale* de la théorie. Lorsqu'on imagine ce que cela donne dans la pratique, on commence à sentir le goût étrange de la théorie des quanta. En effet, si l'on (le physicien-observateur) choisit de mesurer la position d'une particule avec une certaine précision, le fait de monter l'expérience et d'effectuer la mesure interdit à tout jamais, et de façon irréversible, de mesurer la quantité de mouvement avec une précision comparable. D'après la théorie, cependant, il n'existe aucun biais avant la mesure : la fonction d'onde ψ est une combinaison linéaire de fonctions qui

décrivent toutes les issues possibles de la mesure, et ce n'est que lorsque la mesure est faite que la fonction d'onde « s'effondre » ou « se projette sur » l'une des issues possibles.

Comme l'a fait remarquer von Neumann, l'instrument de mesure macroscopique est, lui aussi, décrit par une fonction d'onde quantique (mais en fait, dans la pratique, nous n'avons pas besoin de la théorie quantique pour décrire physiquement de tels objets). Von Neumann montra alors formellement qu'il est impossible de relier directement la fonction d'onde de la particule à l'acte de l'observateur de manière à établir la valeur de ψ à toute échelle. L'« effondrement de la fonction d'onde » n'est déterminé qu'au moment précis où l'appareil de mesure et la particule interagissent pour donner une mesure explicite. Wigner attribua cet effondrement à l'intervention de la conscience de l'observateur. Après tout, c'est l'observateur qui décide d'installer le dispositif de mesure, qui choisit de s'intéresser soit à la position, soit à la quantité de mouvement, et qui réalise effectivement la mesure ! Selon le point de vue de von Neumann, pour déterminer l'état de ce dispositif de mesure, il en faut un autre, et cet autre a encore besoin d'un autre, et ainsi de suite à l'infini. Selon le schéma de Wigner, un phénomène ne devient réel (autrement dit, la suite infinie s'interrompt) que lorsque l'observateur en devient conscient.

Pour être tout à fait équitable, il faut dire que d'autres éminents physiciens ont interprété le problème de la mesure quantique sans faire intervenir la conscience de l'observateur. Niels Bohr, le père de la théorie quantique, déclara qu'il n'existe pas de réalité ultime ou profonde ; on ne fait qu'appliquer le principe de complémentarité (dont le principe d'Heisenberg est peut-être l'expression la plus élégante), puis on obtient le résultat dicté par *l'ensemble* de la situation : la mesure, la particule, le dispositif et l'observateur. L'« Interprétation de Copenhague » due à Bohr correspond au point de vue adopté par la plupart des physiciens qui utilisent la théorie. Elle donne une formule décrivant ce qu'on observe avec un dispositif qui, en dernière analyse, est composé du même type de particules quantiques que celles que l'on est en train de mesurer. D'autres physiciens encore ont même suggéré qu'il n'y a pas d'« effondrement » de la fonction d'onde. En revanche, il y aurait de « nombreux mondes », la fonction d'onde prenant, dans chacun d'eux, l'une des valeurs possibles, et la valeur prise dans ce monde étant celle qu'elle prend avec l'observateur que nous voyons ici et maintenant. D'autres encore ont suggéré qu'il existe un « potentiel quantique » qui pour-

rait même faire appel à des signaux se déplaçant plus rapidement que la lumière, ce qui contredit la théorie de la relativité d'Einstein !

J'ai abordé une fois ce problème, au cours d'un déjeuner, avec mon ami Isidor Rabi – un grand physicien – juste quelque cinq mois avant sa mort. Il me regarda en esquissant un sourire mutin et me dit : « La mécanique quantique n'est qu'un algorithme. Utilise-le. Ça marche, ne t'inquiète pas. » Pour le narguer, je lui dis : « Rab, ne me dis pas que tu deviens comme Einstein, que tu commences à avoir des doutes sur toute cette affaire. » Il me répondit en riant : « Écoute, puisque j'ai quelques problèmes avec Dieu en ce moment, pourquoi est-ce que je n'en aurais pas aussi avec la mécanique quantique ? »

Cela nous amène à la question qui nous intéresse : puisque la théorie est déjà si étrange, pourquoi ne pas la rendre encore plus étrange en imaginant qu'il existe des champs ou des dimensions physiques supplémentaires, non encore découverts, qui pourraient nous révéler la vraie nature de la conscience ? En fait, il s'agit là d'une façon subtile, mais également plus repoussante, de faire encore une fois de la physique un épouvantail de service. Le point de vue adopté par le mathématicien et astrophysicien Roger Penrose dans son livre de grande envergure intitulé *The Emperor's New Mind*, qui a justement pour thème la nature de la conscience, en est un bon exemple. Le livre abonde en exemples très clairs des paradoxes de la physique et en descriptions des limitations axiomatiques des mathématiques. En se fondant sur sa seule intuition, issue de son expérience personnelle de mathématicien et de l'appréciation de ces limites axiomatiques, Penrose rejette l'idée selon laquelle le cerveau serait un ordinateur. Il met en évidence les limites de la mécanique quantique et de la relativité dans des domaines où les dimensions sont si petites (inférieures à ce qu'on appelle la longueur de Planck, qui est égale à 10^{-33} cm) que de telles théories ne s'appliquent plus. Et il en appelle à une théorie de la gravitation quantique qui permettrait d'étendre ces théories. Ensuite, grâce à une remarquable pirouette, il suggère que le mystère de la conscience sera résolu lorsqu'une théorie satisfaisante de la gravitation quantique aura été construite.

Je suppose que son raisonnement est le suivant : les décisions de l'observateur et de l'opérateur *font* partie intégrante des mesures quantiques et relativistes. L'esprit de l'observateur construit et applique des théories mathématiques dont les énoncés transcendent la capacité de confirmation et d'infirmation par l'axiomatique for-

melle. Néanmoins, contrairement à un ordinateur, l'observateur peut en vérifier la véracité et le sens. Comme tout le reste, le cerveau de l'observateur, et en particulier ses synapses – là où se passent la plupart des choses –, sont composés, en dernière analyse, de particules qui obéissent aux lois quantiques. Les lois physiques telles qu'elles sont formulées actuellement ne rendent pas compte de la conscience. Et elles sont également incapables d'expliquer la gravitation quantique. Mais alors, peut-être que si nous parvenions à expliquer la gravitation quantique, elle nous fournirait des pistes à propos de la conscience, qui semble planer au-dessus de toutes nos théories !

Il est clair que la physique tient ici le rôle d'épouvantail de service – plus raisonné, peut-être, que nombre de ceux dont nous parlent les tracts religieux et les livres d'occultisme, mais pas plus bénéfique pour autant. En fait, bien que le livre de Penrose contienne un grand nombre d'excellentes descriptions physiques, il a peu de chose à voir avec le problème de la conscience en tant qu'intentionnalité, car il néglige les connaissances à la fois psychologiques et biologiques qui sont essentielles à la compréhension du problème. La description de Penrose est un peu comme celle d'un écolier qui, au cours d'un examen, ne sachant pas la formule de l'acide sulfurique, fait à la place une merveilleuse description de son chien Médor.

Ce qui manque à la description de Penrose et à d'autres, c'est une sobre analyse scientifique des structures et des fonctions les plus liées à l'état conscient : une description de la vraie psychologie, des vrais cerveaux et de leur biologie sous-jacente. Bien que la physique fournisse évidemment les bases nécessaires à la biologie, elle ne s'intéresse ni aux structures, ni aux processus, ni aux principes biologiques. Ceux-ci sont très particuliers et bien plus manifestement liés à l'esprit que les idées générales de symétrie et de mesure quantique – aussi essentielles ces idées soient-elles à la compréhension de l'existence de toute chose. En fait, il est beaucoup plus raisonnable de construire et de tester une théorie de l'esprit fondée sur des processus biologiques que de postuler que des théories physiques exotiques pourraient en fournir une explication. Après tout, il existe suffisamment de preuves directes du fait que l'anatomie influe sur la conscience.

Par conséquent, tant que nous n'aurons pas atteint une impasse biologique, nous ferions bien de rejeter, en déclarant qu'il s'agit là d'une erreur de catégorie, l'idée que la physique exotique elle-même

puisse être capable de fournir une description de la conscience de l'observateur. Nous ne devons pas confondre les bases de nos mécanismes mentaux avec nos élégantes constructions intellectuelles, dont les théories physiques sont un exemple. (Voici une description un peu crue d'un concours hippique qui pourra faire saisir, du moins à ceux qui comprennent l'anglais, ce qu'est une erreur de catégorie : « *a horse show is a bunch of horses showing their asses to a bunch of horses' asses who are showing their horses* [2] ».)

Nous pouvons au moins remercier Penrose – un grand esprit scientifique – d'avoir à nouveau attiré l'attention sur une erreur de catégorie encore plus courante : celle qui consiste à assimiler le cerveau à un ordinateur. Tournons-nous à présent vers cette question, car elle nous permettra de nous rapprocher beaucoup plus de la question centrale que ne le feraient des considérations physiques supplémentaires.

ORDINATEURS NUMÉRIQUES : LA FAUSSE ANALOGIE

Si le fait d'utiliser la physique comme épouvantail de service ne mène à rien, que dire de cet étonnant objet ou construction physique qu'est l'ordinateur numérique ? Après tout, l'ordinateur – la plus remarquable invention du XXe siècle – semble s'acquitter d'un nombre tout aussi remarquable de fonctions qui, à première vue, ressemblent à des fonctions mentales.

Des choses extraordinairement absurdes ont été dites à propos de l'idée selon laquelle les machines seraient capables de penser. Pour la plupart, l'absurdité provient de l'analogie entre pensée et logique. Il est un fait indiscutable : les ordinateurs effectuent des opérations logiques. Mais le hic est que les opérations logiques à elles seules, telles qu'elles sont utilisées par les ordinateurs, ressemblent aussi peu à la pensée que les phénomènes physiques permettant d'additionner des nombres à l'aide d'un boulier ne ressemblent à ce qui se passe dans le cerveau d'un mathématicien lorsqu'il fait, ou qu'il crée, de l'arithmétique.

Pour comprendre l'origine de cette confusion, je dois explorer quelque peu la théorie qui se trouve derrière les ordinateurs numé-

2. « Un concours hippique c'est une bande de chevaux exhibant leur cul à une bande de crétins (littéralement " culs de chevaux ") qui exhibent leurs chevaux. » (N.d.T.)

riques. Cette théorie est due en grande partie au travail d'Alan Turing, mathématicien britannique qui se suicida en croquant une pomme empoisonnée. La justice britannique ayant découvert que Turing était homosexuel, elle l'avait obligé à choisir entre la prison et la « castration chimique » par l'intermédiaire d'un traitement à base d'œstrogènes (des hormones féminines). Turing choisit le traitement hormonal, ce qui entraîna des effets féminisants sur son organisme – et qui sait quels effets sur son cerveau. Quoi qu'il en soit, ce cerveau-là donna naissance à un puissant ensemble d'idées mathématiques, l'une d'elles étant connue sous le nom de machine de Turing.

Turing définit une classe d'automates abstraits et montra que tous les membres de cette classe étaient capables de calculer tous les éléments d'une vaste classe de fonctions. (À l'exception de quelques machines déterminées, tous les ordinateurs sont des machines de Turing.) Une machine de Turing (figure P-2) est une machine qui possède un nombre fini d'états et un ruban, ou bande, infini, divisé en cases (contenant chacune un 0 ou un 1) ; dans chaque case de la bande, la machine peut écrire soit un 0, soit un 1, et elle peut faire glisser la bande d'une case vers la gauche ou vers la droite. La machine possède des instructions qui contiennent des conditions et des actions, et elle effectue une action lorsqu'une condition particulière est satisfaite. La condition est déterminée par le symbole qui se trouve sur la bande, sous la tête de lecture/écriture, et par l'état de la machine. Les actions possibles se limitent aux quatre décrites ci-dessus, après quoi la machine passe à l'état suivant spécifié par le programme. Une « machine de Turing universelle » est une machine de Turing capable de simuler n'importe quelle machine de Turing particulière. (Les machines de Turing particulières peuvent avoir des mécanismes et des parties différentes, pourvu qu'elles obéissent à la description de Turing.)

Nous voici arrivés à un stade où nous allons être tentés de commettre une erreur de catégorie. En effet, il existe tout un corps d'arguments très convaincants qui stipulent que si je peux décrire une procédure mathématique effective (le terme technique est algorithme ; voir la figure P-4), alors cette procédure pourra être exécutée par une machine de Turing. Plus généralement, nous savons que *tout* algorithme, ou procédure effective, peut être exécuté par n'importe quelle machine de Turing universelle. Le fait qu'il existe des machines universelles incite à penser que le *mécanisme* par

lequel elles opèrent est sans importance. On peut montrer que cela est vrai dans le monde réel en faisant s'exécuter un programme donné sur deux ordinateurs numériques aux architectures, ou conceptions matérielles, totalement différentes, et en constatant que les deux parviennent au même résultat (voir la figure P-3).

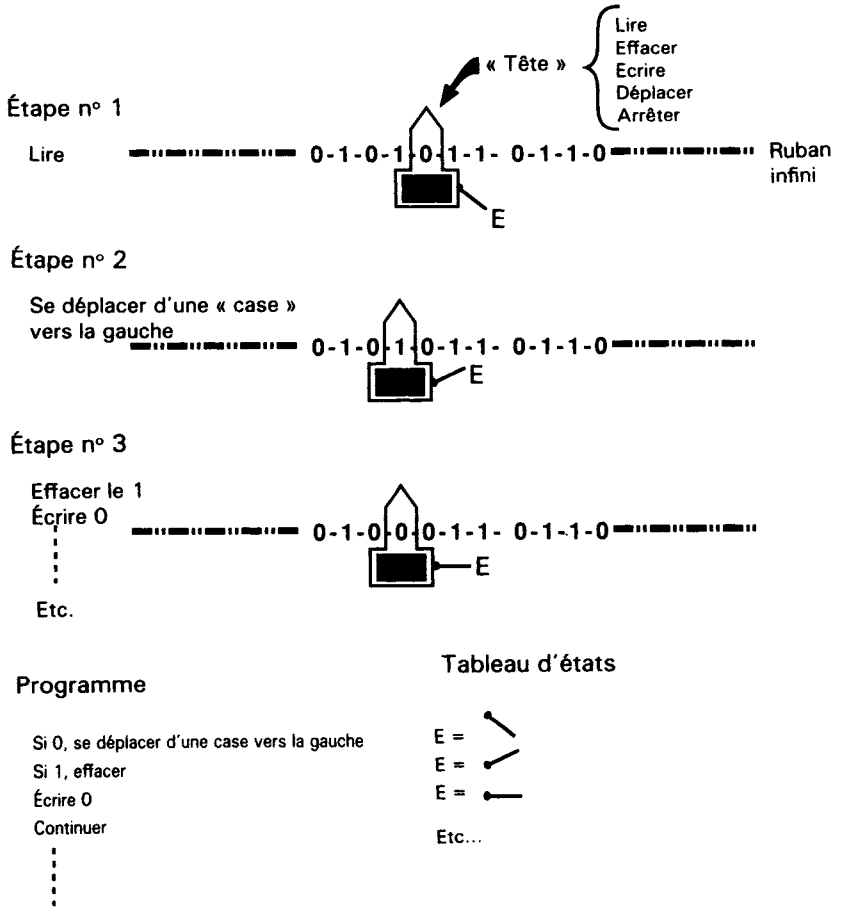


FIGURE P-2

Une machine de Turing. Il a été démontré que cet objet abstrait représente les opérations fonctionnelles de pratiquement tous les ordinateurs. L'analyse de Turing s'applique aux ordinateurs réels, même s'il est vrai qu'une machine de Turing (contrairement à un vrai ordinateur) serait obligée de passer par un nombre d'étapes beaucoup plus élevé qu'il n'en faudrait pour exécuter en un temps raisonnable un traitement de l'information ou algorithme simple . La notion de machine de Turing constitue un exploit de clarté théorique.

En se fondant sur ces propriétés, on a considéré que les mécanismes cérébraux résultaient d'un processus « fonctionnel », dont on affirmait qu'il pouvait être décrit à la manière d'un algorithme. Ce point de vue est celui du fonctionnalisme (l'une de ses formes les plus radicales étant justement le fonctionnalisme des machines de Turing). Le fonctionnalisme suppose que la psychologie peut être décrite de manière adéquate d'après l'« organisation fonctionnelle du cerveau » – tout comme, en informatique, les performances du matériel sont déterminées par le logiciel. Le fonctionnalisme s'intéresse non seulement aux fonctions assurées par divers systèmes, mais aussi aux relations existant entre leurs composants, notamment dans la mesure où elles donnent lieu à d'autres relations. Les théories fonctionnalistes sont indifférentes à la mise en œuvre mécanique des systèmes, et elles traitent donc ces relations en des termes abstraits.

Selon le point de vue fonctionnaliste, seuls les algorithmes importent, en dernière analyse, pour comprendre la psychologie – et non le matériel sur lequel ces algorithmes sont exécutés. D'après le fonctionnalisme, il est donc possible de décrire correctement à l'aide d'algorithmes ce que le cerveau fait. De plus, l'organisation et la composition des tissus cérébraux ne sont pas importantes du moment que l'algorithme « marche », c'est-à-dire qu'il parvient au résultat puis s'arrête (figure P-4). (Cette position « libérale », selon laquelle la présence de tissus cérébraux particuliers n'est pas nécessaire, a d'ailleurs envahi une grande partie de la psychologie cognitive actuelle.)

Si nous acceptons ce point de vue, une analyse issue de la logique formelle (connue sous le nom de thèse de Church) nous incite à penser que s'il existe une méthode de calcul qui immanquablement parvienne à résoudre un problème donné puis s'arrête, alors il existe une méthode susceptible d'être exécutée par une machine de Turing et d'aboutir exactement au même résultat. Autrement dit, pour tous les problèmes pouvant être systématiquement résolus au bout d'un temps fini donné, les machines de Turing sont aussi puissantes que n'importe quelle autre entité de résolution, *y compris le cerveau.* Donc, d'après cette analyse, soit le cerveau est un ordinateur, soit l'ordinateur est un analogue du cerveau, ce qui signifie que les ordinateurs constituent un modèle adéquat pour tout ce que le cerveau fait d'intéressant.

Ce genre d'analyse sous-tend ce qui est aujourd'hui connu sous le nom d'hypothèse des systèmes symboliques physiques, et qui est à la base de la majeure partie des recherches en intelligence

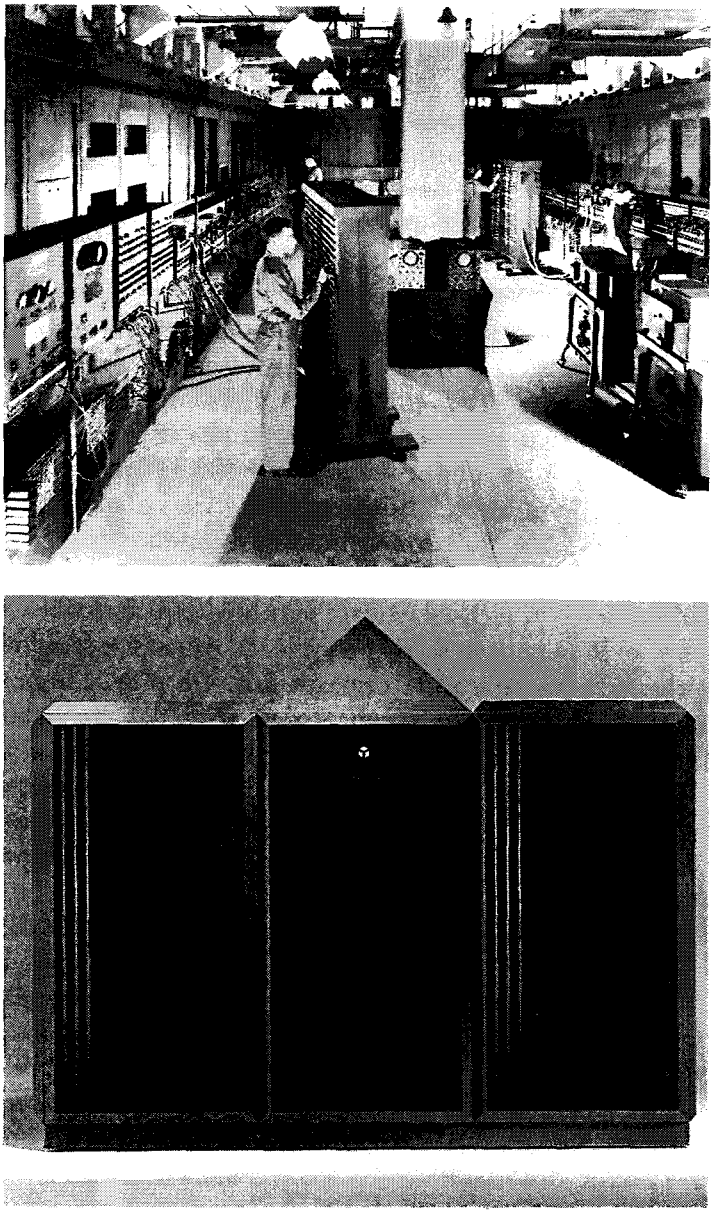
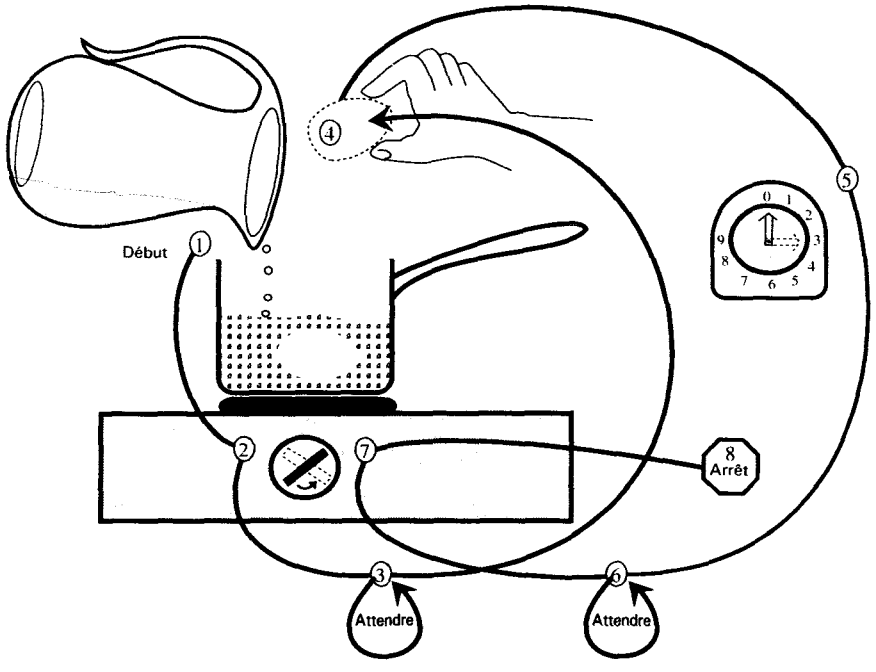


FIGURE P-3

Deux ordinateurs réels. En haut *: l'*ENIAC, *qui fut le premier ordinateur numérique utilisable.* En bas *: un* N-CUBE, *ordinateur disponible sur le marché, fondé sur le traitement massivement parallèle de l'information. L'*ENIAC *remplissait une grande pièce et exécutait environ cinq mille instructions par seconde ; le* N-CUBE *a à peu près la taille d'une table de bureau normale et il est capable d'exécuter près de huit milliards d'instructions par seconde. Si vous avez les moyens de vous en payer un, il réduira votre temps de calcul, mais son principe de base (celui de Turing) est le même que celui des autres ordinateurs.*

artificielle. Cette hypothèse stipule que les fonctions cognitives sont assurées par la manipulation de symboles suivant certaines règles. Dans les systèmes symboliques physiques, les symboles sont représentés dans les programmes par les états d'objets physiques. Des



ALGORITHME PERMETTANT DE FAIRE CUIRE UN ŒUF

① Mettre de l'eau dans une casserole

② Allumer la plaque électrique

③ Si l'eau ne bout pas, aller à l'étape ③ sinon passer à ④

④ Mettre l'œuf dans l'eau

⑤ Régler le minuteur sur trois minutes

⑥ Si le minuteur n'a pas sonné, aller à l'étape ⑥ sinon passer à l'étape ⑦

⑦ Éteindre la plaque et refroidir

⑧ C'est fini... sortir l'œuf, l'éplucher et le manger

FIGURE P-4

Algorithme permettant de faire cuire un œuf. L'algorithme permettant d'additionner deux nombres comporterait des instructions tout aussi explicites.

chaînes de symboles sont utilisées pour représenter les entrées sensorielles, les catégories, les comportements, la mémoire, les propositions logiques et de fait toute l'information traitée par le système. Les opérations nécessaires à la transformation des chaînes de symboles d'entrée en chaînes de symboles de sortie sont des

calculs et, selon l'hypothèse des systèmes symboliques physiques, elles peuvent donc être effectuées par n'importe quelle machine de Turing dûment programmée. Comme je l'ai déjà dit, ces opérations sont de nature purement formelle ; autrement dit, elles peuvent être exécutées en l'absence de toute référence à la *signification* des symboles concernés. (Comme nous l'avons vu au cours des chapitres 2 et 12, un tel ensemble de règles constitue ce qu'on appelle une syntaxe.) L'architecture particulière du dispositif de calcul qui fonctionne selon ces règles syntaxiques n'est importante que dans la mesure où elle doit satisfaire certaines exigences de rapidité et de capacité de mémoire afin d'être capable d'achever son travail en un temps raisonnable.

Pourquoi est-ce que ce point de vue n'est pas satisfaisant ? Les raisons sont multiples, mais avant de les aborder, je voudrais vous rappeler qu'il existe un rapport entre les systèmes symboliques physiques et l'argument en faveur du fonctionnalisme (qui possède de nombreuses variantes, ayant toutes en commun la notion de cause formelle). En effet, si l'une quelconque des formes du fonctionnalisme était une bonne théorie de l'esprit, alors le cerveau serait véritablement analogue à une machine de Turing. Et, dans ce cas, la description pertinente tant de l'un que de l'autre se situerait au niveau des représentations symboliques et des algorithmes, et non au niveau biologique.

Toutes les formes de la théorie fonctionnaliste n'imposent pas un si grand degré d'identification des processus mentaux avec les processus des machines de Turing. La position la plus tranchée, initialement formulée par Hilary Putnam, et connue sous le nom de « fonctionnalisme des machines de Turing », postule que les deux types de processus sont totalement équivalents. Aujourd'hui, cette vision n'est plus largement admise ; en fait, elle a été rejetée par Putnam lui-même. Quant aux formes plus faibles du fonctionnalisme, elles n'exigent pas une équivalence stricte entre les états cérébraux et ceux des machines de Turing. Cependant, *toutes* les formes de fonctionnalisme stipulent que deux systèmes ayant des états fonctionnels isomorphes se trouvent nécessairement dans des états cognitifs identiques, quelles que soient leurs différences de constitution physique. Cette conclusion est un proche cousin de certains résultats de Turing sur le calcul universel. Ces résultats peuvent être résumés en disant que deux ordinateurs dont les tableaux abstraits de transition d'état sont identiques et qui portent des symboles identiques sur leurs bandes respectives (voir la figure

P-2 pour les définitions et des exemples) sont en train d'effectuer les mêmes calculs, quelle que soit la forme physique prise par le processeur et la bande de chacun.

Et maintenant, le coup de grâce (en fait, ils sont multiples) ! L'analyse de l'évolution, du développement et de la structure du cerveau rend hautement improbable l'idée que le cerveau puisse être une machine de Turing. Comme nous l'avons vu au chapitre 3, le cerveau fait preuve d'une énorme variabilité structurelle entre individus à de nombreux niveaux d'organisation ; un examen de la façon dont il se développe indique que le cerveau est une structure hautement variable. Et en fait, un simple calcul montre que le génome d'un être humain (l'ensemble des gènes d'un individu) ne suffit pas à spécifier explicitement la structure synaptique du cerveau en cours de développement. En outre, le comportement de chaque organisme est biologiquement individualisé et extrêmement diversifié, et cela indépendamment du fait que l'organisme soit ou non capable, comme les êtres humains, d'enregistrer et de décrire des expériences subjectives.

Plus mortel encore est le fait que l'analyse de la variabilité écologique, environnementale, et celle des procédures de catégorisation des animaux et des humains (dont je parlerai dans la section suivante) font qu'il est improbable que le monde (physique et social) puisse fonctionner comme la bande d'une machine de Turing. Un raisonnement semblable a amené Putnam à répudier son modèle fonctionnaliste originel et d'autres modèles dérivés. Son argument central est que les états psychologiques, et notamment les attitudes propositionnelles (« croire que *p* », « souhaiter que *p* », etc.), ne peuvent être décrits par le modèle computationnel. Nous ne pouvons individualiser des concepts et des croyances sans faire référence à un environnement. Le cerveau et le système nerveux ne peuvent pas être considérés indépendamment des états du monde et des interactions sociales. Mais de tels états, qu'ils soient environnementaux ou sociaux, sont indéterminés et non figés. Ils ne peuvent pas être simplement identifiés par une quelconque description logicielle. Et le fonctionnalisme, interprété dans ce contexte comme l'idée que les attitudes propositionnelles sont équivalentes à des états computationnels du cerveau, n'est donc pas tenable.

John Searle, un autre philosophe, a également critiqué fortement la conception fonctionnaliste. Son désaccord est fondé sur l'idée selon laquelle il n'existe aucune spécification purement computationnelle qui fournisse des conditions suffisantes à l'apparition de

la pensée ou des états intentionnels. Son argument (qui s'applique à la conscience d'ordre supérieur, le type de conscience dont les humains sont dotés) consiste à dire que les programmes d'ordinateur sont strictement définis par leur structure syntaxique formelle, que la syntaxe ne suffit pas pour obtenir des capacités sémantiques et que, en revanche, l'esprit humain est *caractérisé* par le fait qu'il possède un contenu sémantique. Le contenu sémantique fait appel à des significations, alors que la syntaxe proprement dite ne s'intéresse pas aux significations. Dans cette perspective, le rejet du fonctionnalisme devient inévitable. De plus, Searle affirme que, dans la mesure où la conscience correspond chez les humains à un type d'intentionnalité qui s'accompagne inévitablement d'expériences subjectives, il s'ensuit que, par définition, aucun organisme ne peut avoir des états intentionnels s'il lui manque l'expérience subjective. Or les ordinateurs ne possèdent pas ce genre d'expérience. D'ailleurs, certains fonctionnalistes (probablement la majorité d'entre eux) limitent leurs affirmations à des énoncés qui excluent les propriétés subjectives ou phénoménales. Mais étant donné le genre d'arguments que Searle utilise, il rejetterait quand même leurs affirmations (et à mon avis il aurait raison de le faire), car elles n'ont aucun rapport avec l'origine de la conscience et de la pensée.

C'est la notion de signification qui est en jeu ici. La signification, comme le dit Putnam, « est interactive. L'environnement lui-même contribue à déterminer ce à quoi les mots d'un locuteur, ou d'une communauté, font référence ». Et comme cet environnement n'est pas figé, il n'admet aucune description générale *a priori* en termes de procédures effectives. De plus, nous avons vu au cours de ce livre que le propre corps du locuteur joue un rôle tout aussi déterminant au niveau sémantique. Les arguments portant sur la sémantique et la signification sont importants pour toute théorie de la conscience (et de la pensée) dont les références canoniques sont notre propre expérience phénoménale en tant qu'humains et notre capacité de décrire cette expérience grâce au langage.

Remarquez à présent comme les ordinateurs sont différents. En ce qui concerne les ordinateurs habituels, nous n'avons pas grand mal à accepter le point de vue fonctionnaliste, puisque la seule signification possible des symboles qui se trouvent sur la bande et des états internes du processeur est *la signification qui leur a été attribuée par un programmeur humain.* Il n'y a aucune ambiguïté dans l'interprétation des états physiques sous forme de symboles,

parce que les symboles sont représentés numériquement d'après des règles de syntaxe. Le système est *conçu* pour passer rapidement d'un état bien défini à un autre, en évitant la zone de transition qui les sépare ; du point de vue électrique, chaque composant bascule toujours soit vers « 0 », soit vers « 1 ». Les petites déviations qui surviennent quand même dans les paramètres physiques (les niveaux de bruit, par exemple) ne sont pas prises en compte par convention et conception. D'ailleurs, l'un des objectifs de toutes ces conventions est de garantir que toute différence existant entre deux systèmes en raison des différences dans leur manière de représenter physiquement les symboles n'aura effectivement aucune signification. Les différences au niveau matériel ne posent aucun problème pourvu que le matériel fonctionne correctement. Rappelez-vous, cependant, que cette portabilité des systèmes fonctionnalistes d'une mise en œuvre matérielle à une autre a un prix, qui est que les processus fonctionnels primitifs doivent nécessairement opérer sur des représentations symboliques de l'information.

Nous commençons à présent à voir pourquoi les ordinateurs numériques sont de faux analogues du cerveau. L'analogie facile avec les ordinateurs numériques s'effondre pour différentes raisons. Tout d'abord, la bande lue par les machines de Turing porte des symboles non ambigus choisis dans un ensemble fini ; en revanche, les signaux sensoriels qui sont à la disposition du système nerveux sont de nature authentiquement analogique et, par conséquent, ils ne sont ni dépourvus d'ambiguïté ni en nombre fini. Ensuite, par définition, les machines de Turing possèdent un nombre fini d'états internes, tandis que le nombre d'états pouvant être pris par le système nerveux humain (par exemple, par modulation analogique de grands nombres de liaisons synaptiques dans les connexions neuronales) semble illimité. Par ailleurs, les transitions entre états des machines de Turing sont totalement déterministes, alors que celles des humains semblent au contraire largement indéterminées. Enfin, l'expérience humaine ne se fonde pas sur une abstraction aussi simple qu'une machine de Turing ; pour former nos « significations », nous devons grandir et communiquer au sein d'une société.

La beauté abstraite des machines de Turing est séduisante. Mais, même en science, où elles confèrent habituellement une grande puissance à notre pensée, on doit se méfier des abstractions excessives. Dans certains contextes, l'abstraction devient absurde. Cela me rappelle une histoire à propos d'un champ de courses qui perdait

de l'argent. La direction du champ de courses décida de consulter trois experts : un comptable, un ingénieur et un physicien. Le comptable leur conseilla de restructurer le bilan ; l'ingénieur leur suggéra de mieux aménager le terrain de courses, en surélevant les virages et en améliorant le drainage. Quand ce fut au tour du physicien, celui-ci se dirigea vers le tableau, dessina un cercle et dit : « Supposons que le cheval soit une sphère. »

Contrairement à ce qui se passe avec les ordinateurs, les caractéristiques des réponses du système nerveux dépendent de l'histoire individuelle de chaque système, parce que ce n'est qu'*à travers les interactions avec le monde* qu'il est possible de sélectionner des réponses adéquates. Et la diversité des expériences vécues par chaque individu introduit des variations non seulement entre les différents systèmes nerveux, mais aussi au cours du temps pour un même système. L'existence d'une grande variabilité au niveau des individus dans les systèmes cognitifs (voir le chapitre 3) va à l'encontre du postulat de base du fonctionnalisme, selon lequel les représentations possèdent une signification indépendamment de leur mise en œuvre physique. Ainsi, il apparaît que pour parvenir à un niveau non trivial de performance cognitive, il faille abandonner cette caractéristique si chère aux systèmes fonctionnalistes qu'est l'indépendance par rapport à la mise en œuvre physique. (Cela ne signifie pas qu'ayant abandonné le point de vue libéral du fonctionnalisme, nous devions nécessairement basculer vers la position chauviniste qui se trouve à l'extrême opposé : l'idée que la chimie du carbone, les tissus humides, etc., sont *absolument* nécessaires à la cognition. Si tel était le cas, les objets dont j'ai parlé au chapitre 19 ne pourraient pas être construits.)

Quel que soit le type de représentations internes utilisé par les systèmes fonctionnalistes, il faut une procédure pour définir, dans ces représentations, la signification des unités individuelles (les symboles ou leurs généralisations) ainsi que celle de leurs combinaisons. Mais on a du mal à voir comment, en l'absence d'un programmeur, un mécanisme capable d'attribuer une signification aux représentations syntaxiques pourrait être construit qui préserverait le caractère arbitraire de ces représentations, caractère essentiel à la position fonctionnaliste. Or c'est là précisément la situation poignante dans laquelle nous nous trouvons : nous n'avons pas de programmeur, pas d'homoncule dans la tête.

Je ne pourrais pas conclure ici sans évoquer le fait que, au cours des dernières années, une grande masse de travaux a été accomplie

sur les modèles « connexionnistes » des processus perceptifs et cognitifs, ou « réseaux neuronaux ». Il s'agit là de modèles formels dans lesquels les connexions entre éléments du réseau subissent des modifications qui sont grosso modo analogues à celles qui surviennent dans les synapses. Je suppose que cela justifie la métaphore concernant le « neuronal », mais à d'autres égards cette métaphore devient exagérée, comme nous allons le voir dans ce qui suit.

Ces constructions ont été utiles dans un certain nombre d'applications. Un grand nombre de ces modèles partent d'hypothèses concernant la nature des systèmes intelligents qui sont semblables à celles émises par les spécialistes de l'intelligence artificielle. Mais, contrairement aux travaux classiques en intelligence artificielle, ces modèles utilisent des processus distribués sur des réseaux, et les modifications dans les connexions se passent en partie sans l'intervention d'une programmation précise. Néanmoins, les systèmes connexionistes exigent qu'un programmeur ou un opérateur spécifient leurs entrées et leurs sorties, et ils font appel à des algorithmes pour l'introduction de ces spécifications. Et, bien que ces systèmes autorisent des modifications résultant de l'« expérience acquise », un tel mécanisme d'« apprentissage » dérive d'un ensemble d'instructions et non d'une sélection. Dans les systèmes connexionnistes – contrairement à ce qui se passe dans les systèmes sélectifs, qui effectuent des catégorisations fondées sur des valeurs – ce sont les *réponses* (et non pas les valeurs) qui sont spécifiées à l'avance et imposées au système par un opérateur humain, dans des conditions adéquates et en présence d'un mécanisme rétroactif de correction d'erreurs, afin de permettre l'apprentissage.

Les architectures des réseaux neuronaux ne correspondent pas à la réalité biologique, et ces réseaux « fonctionnent » d'une façon tout à fait différente de celle du système nerveux. Ils utilisent des matrices de connexions denses et symétriques. En général, ils ne ressemblent pas du tout aux structures neuronales et à l'anatomie que j'ai décrites au cours de ce livre. Si l'on adoptait les réseaux neuronaux comme modèle standard de la structure et de la fonction du cerveau, on serait obligé de dire qu'ils prêchent en faveur de la vision du cerveau en tant que machine de Turing. Ainsi, quels que soient leur intérêt et leur utilité, les réseaux neuronaux ne sont pas des modèles adéquats ou des analogues de la structure cérébrale. (Les lecteurs qui voudront approfondir ces questions trouveront des références à deux recueils d'articles sur ce thème dans la bibliographie, à la fin du livre.)

Que l'on se fonde sur les ordinateurs numériques ou sur les modèles connexionnistes, on se retrouve devant un même problème embarrassant. En effet, lorsqu'on considère que le cerveau est une machine de Turing, on est confronté à plusieurs constatations troublantes, à savoir que l'hypothétique tableau des états et des transitions d'état du cerveau (voir la figure P-2) n'est pas connu, que les symboles inscrits sur la bande d'entrée sont ambigus et n'ont pas de signification fixée à l'avance, et que les règles de transition, quelles qu'elles soient, ne sont pas appliquées de façon cohérente. De plus, chez les animaux du monde réel, les entrées et les sorties ne sont pas spécifiées par un entraîneur ou un programmeur. Il semble donc bien que l'on ne gagne rien ou presque rien à appliquer cette analogie ratée entre ordinateur et cerveau.

Cependant, il n'est pas si facile d'abandonner ce domaine une fois pour toutes. Il existe une masse de travaux de psychologie cognitive fondés sur des malentendus du même type concernant les hypothèses que l'on peut émettre sur le fonctionnement du cerveau sans se préoccuper d'analyser comment il est construit du point de vue physique. Tournons-nous donc à présent vers quelques-uns des problèmes soulevés, en psychologie cognitive, par l'un des concepts centraux de cette discipline : le concept de représentation mentale.

QUELQUES CERCLES VICIEUX DANS LE PAYSAGE COGNITIF

Les sciences cognitives – un mélange de psychologie, d'informatique, de linguistique et de philosophie – ont pris énormément d'importance. Comme c'est le cas chaque fois que de sérieux efforts sont faits dans un domaine – avec ou sans fondements –, il en est sorti beaucoup de choses très intéressantes, à la fois pour les scientifiques et les non-scientifiques. L'un de ces résultats positifs – et non des moindres – a été la mise à l'écart du behaviorisme simpliste. Mais en même temps, une extraordinaire méprise sur la nature de la pensée, du raisonnement, de la signification et de leur rapport avec la perception a pris forme qui menace de miner à la base l'ensemble de l'entreprise.

Remonter aux origines de ce malentendu demande un certain effort, car il possède des racines historiques, intellectuelles et pratiques complexes. Je dois d'ailleurs vous avertir que nous allons pour cela être amenés à approfondir un certain nombre de questions complexes dont je ne pourrai pas simplifier la description au-delà d'un certain point. Mais, avant d'aborder les choses dans le détail,

je voudrais vous donner un bref aperçu du malentendu en question. Il est issu de l'idée selon laquelle les objets qui peuplent le monde appartiennent à des catégories fixes, que les choses possèdent des descriptions essentielles, que les concepts et le langage s'appuient sur des règles qui acquièrent un sens par attribution formelle à des catégories fixes du monde, et que l'esprit opère à travers ce qu'on appelle des représentations mentales. Selon certains, ces représentations sont censées s'exprimer dans un langage de la pensée – ou « mentalais », selon le terme utilisé par le philosophe Jerry Fodor. Attribuer un sens consiste à faire correspondre *exactement* les symboles d'un tel langage aux entités ou catégories du monde définies par des conditions nécessaires et suffisantes (les catégories classiques). Ainsi, la spécification des règles servant à manipuler les représentations (qui constituent une syntaxe), si elle est complète, peut être réalisée par un dispositif de calcul. Le cerveau, d'après ce point de vue, est une sorte d'ordinateur. (Remarquez le parallèle entre certaines de ces affirmations et celles de la section précédente.)

L'adoption de ce point de vue sous l'une ou l'autre de ses formes s'est beaucoup répandue en psychologie, en linguistique, en informatique et en intelligence artificielle. Or c'est l'un des plus remarquables malentendus de l'histoire des sciences. En fait, non seulement ce point de vue est-il en désaccord avec les données connues de biologie humaine et de neurobiologie, mais il constitue également une erreur de catégorie extrêmement grave.

Nous nous sommes leurrés en partie à cause des succès que nous avons remportés dans les sciences « dures » en retirant l'esprit de la nature. Nous avons commis l'erreur d'attribuer les caractéristiques des constructions mentales humaines (telles que la logique ou les mathématiques) au raisonnement humain et au monde macroscopique dans lequel nous vivons. Chaque fois que je songe à la façon dont ces cercles vicieux conçus par la raison se sont gravés à la surface du paysage cognitif, je ne peux m'empêcher de penser à cette histoire qu'on raconte à propos d'une conversation entre deux souris dans un laboratoire de psychologie. Ayant réussi à sortir d'un labyrinthe, l'une d'elles dit à l'autre : « Tu sais, je pense que j'ai finalement réussi à entraîner mon psychologue. Chaque fois que je parcours correctement un labyrinthe, il me donne un morceau de fromage. »

Afin de vous montrer pourquoi les notions de « mentalais », de règles et de représentations, et enfin les notions calculatoires ne peuvent pas marcher, je dois aborder quelques-unes des hypothèses

les plus strictes du fonctionnalisme sous-jacent à la psychologie cognitive. Ensuite, je devrai considérer une conception du monde (et en particulier du monde scientifique) dite objectiviste. Enfin, je devrai m'intéresser à une question centrale : les données concernant la façon dont nous catégorisons effectivement le monde, à la fois du point de vue perceptif et conceptuel. Une fois cela fait, nous serons en mesure de voir les erreurs de raisonnement qui menacent de faire échouer l'entreprise cognitiviste. Cependant, comme les arguments et les données dont il sera question ici ne sont pas exhaustifs, j'encourage vivement les lecteurs à consulter les ouvrages correspondants, cités dans la section bibliographique, afin d'en savoir plus. Quant à moi, j'essayerai d'esquisser les questions d'une façon minimale mais incisive, car elles se trouvent au cœur de toute tentative de compréhension de la question de l'esprit.

La plupart de ceux qui travaillent en psychologie cognitive tiennent aux points de vue que je dénonce ici. Mais une minorité d'entre eux ont cependant adopté des points de vue opposés, et à beaucoup d'égards très semblables au mien. Ces penseurs sont issus de nombreux domaines – de la psychologie cognitive, de la linguistique, de la philosophie et des neurosciences. Il s'agit notamment de John Searle, Hilary Putnam, Ruth Garret Millikan, George Lakoff, Ronald Langacker, Alan Gauld, Benny Shanon, Claes von Hofsten, Jerome Bruner et sans doute d'autres encore. Je me plais à les considérer comme faisant partie d'un Club des Réalistes, un groupe de gens épars dont les idées sont largement convergentes et qui espèrent qu'un jour les praticiens de psychologie cognitive les plus en vue, ainsi que les neurobiologistes empiriques les plus arrogants, comprendront enfin qu'ils ont été victimes, sans le savoir, d'une escroquerie intellectuelle. Les idées de cette minorité transparaîtront donc dans ce que j'ai à dire, mais il est clair que ces idées varient d'une personne à l'autre. Par conséquent, je prie vivement les lecteurs de consulter directement les travaux de ces spécialistes afin d'analyser de plus près la diversité de leurs pensées et de leurs interprétations.

Les conceptions fonctionnalistes et la représentation sémantique de la signification

L'idée centrale sous-jacente à la plupart des travaux de psychologie cognitive moderne est celle de représentation mentale. Ces

représentations sont abstraites et symboliques (autrement dit, elles représentent une chose ou une relation), elles se forment d'une façon bien définie et elles suivent des règles qui constituent une syntaxe. Elles sont censées être sémantiquement liées au monde par l'intermédiaire de relations fixes et bien définies, ainsi que par l'attribution sémantique de symboles à des objets appartenant à des catégories classiques. Les représentations sont essentielles à la formation de « modèles internes du monde ».

La notion de modèle interne provient des premières idées émises par K. J. W. Craik et, vues sous cet angle, les représentations internes correspondent aux structures externes du monde. Les représentations sont soit propositionnelles – elles font alors appel à des concepts et aux relations existant entre eux –, soit ce sont des images mentales. Les images sont d'origine perceptive, ce qui dans cette perspective – selon les idées fondatrices et très influentes de David Marr, aujourd'hui décédé – correspond à une forme de calcul. Les calculs qui se déroulent sur les structures mentales sont régis par un système de règles (ou syntaxe) et par les représentations elles-mêmes. L'ensemble du système de représentations forme une *lingua mentis* ou mentalais, un langage de la pensée.

Comment aborde-t-on sous cet angle le problème de l'intentionnalité ? Vraisemblablement en déclarant que le sens naît de la mise en correspondance de structures syntaxiques *régies par des règles* avec des objets ou des relations *fixes* et *définis* du monde réel. Une telle sémantique est exhaustive et bien définie et, jointe à sa syntaxe sous-jacente, elle fournit un cadre à la modélisation de la pensée.

Comment une telle conception fonctionnaliste, computationnelle, de l'esprit, si extrêmement formelle et désincarnée, est-elle apparue ? Comment a-t-on pu accepter une conception aussi abstraite de la connaissance humaine, de la raison et de l'activité mentale ? Avant de critiquer cette conception de l'esprit, regardons de plus près la conception correspondante du monde, qui constitue l'un de ses fondements.

L'objectivisme

Le terme « objectivisme » a été utilisé pour caractériser une vision du monde qui semble, à première vue, irréprochablement scientifique et sensée. (L'une des analyses que je compte reprendre ici est celle de Lakoff ; voir la bibliographie.) L'objectivisme va au-delà

des hypothèses du réalisme scientifique, qui stipulent ce qui suit : (1) qu'il existe un monde réel (qui contient les êtres humains mais ne dépend pas d'eux) ; (2) qu'il existe un lien entre les concepts et ce monde ; (3) que ce lien permet l'acquisition de connaissances stables. L'objectivisme suppose, outre les hypothèses du réalisme scientifique, que le monde possède une structure définie, faite d'entités, de propriétés et de leurs inter-relations (figure P-5). Celles-ci peuvent être définies selon les critères classiques de catégorisation, c'est-à-dire selon les critères qui sont nécessaires et suffisants pour définir chaque catégorie. Le monde est organisé de telle sorte qu'il peut être complètement modélisé à l'aide de ce que les mathématiciens et les logiciens appellent des modèles de la théorie des ensembles. Ce type de modèles, que l'on trouve en logique mathématique, se compose d'entités symboliques isolées ou d'ensembles d'entités symboliques, et de leurs relations. Dans ces modèles, on attribue un sens aux symboles (on les rend sémantiquement significatifs) de façon univoque en supposant qu'ils correspondent à des entités et des catégories du monde réel. Certaines des propriétés catégorielles des choses du monde sont considérées comme essentielles, d'autres comme accidentelles.

Du fait de la correspondance singulière et bien définie existant entre les symboles de la théorie des ensembles et les choses telles qu'elles sont définies par catégorisation classique, on peut, sous cet angle, supposer que les relations logiques entre les choses du monde réel existent *objectivement*. Ainsi, ce système de symboles est censé représenter la réalité, et les représentations mentales seront soit vraies, soit fausses, selon qu'elles reflètent ou non correctement la réalité. Du point de vue objectiviste, c'est cette correspondance avec les choses du monde réel qui donne un sens aux expressions linguistiques ; la signification est fondée sur cette définition de la vérité en fonction de ce qui est « correct » ou « incorrect », et la pensée elle-même n'est rien d'autre qu'une manipulation de symboles.

Ce point de vue peut certainement être défendu à l'extérieur du champ scientifique. De fait, la conception objectiviste semble largement en accord avec le bon sens. Mais lorsqu'on adopte cette position en science, on se rapproche beaucoup de la position galiléenne que nous avons évoquée au chapitre 2. Et dans ce sens, les concepts, les affirmations et les langages humains ne sont valides que lorsqu'ils se limitent à la physique, à la chimie et à certaines parties de la biologie.

Nous allons voir que, aussi sensé que ce point de vue puisse

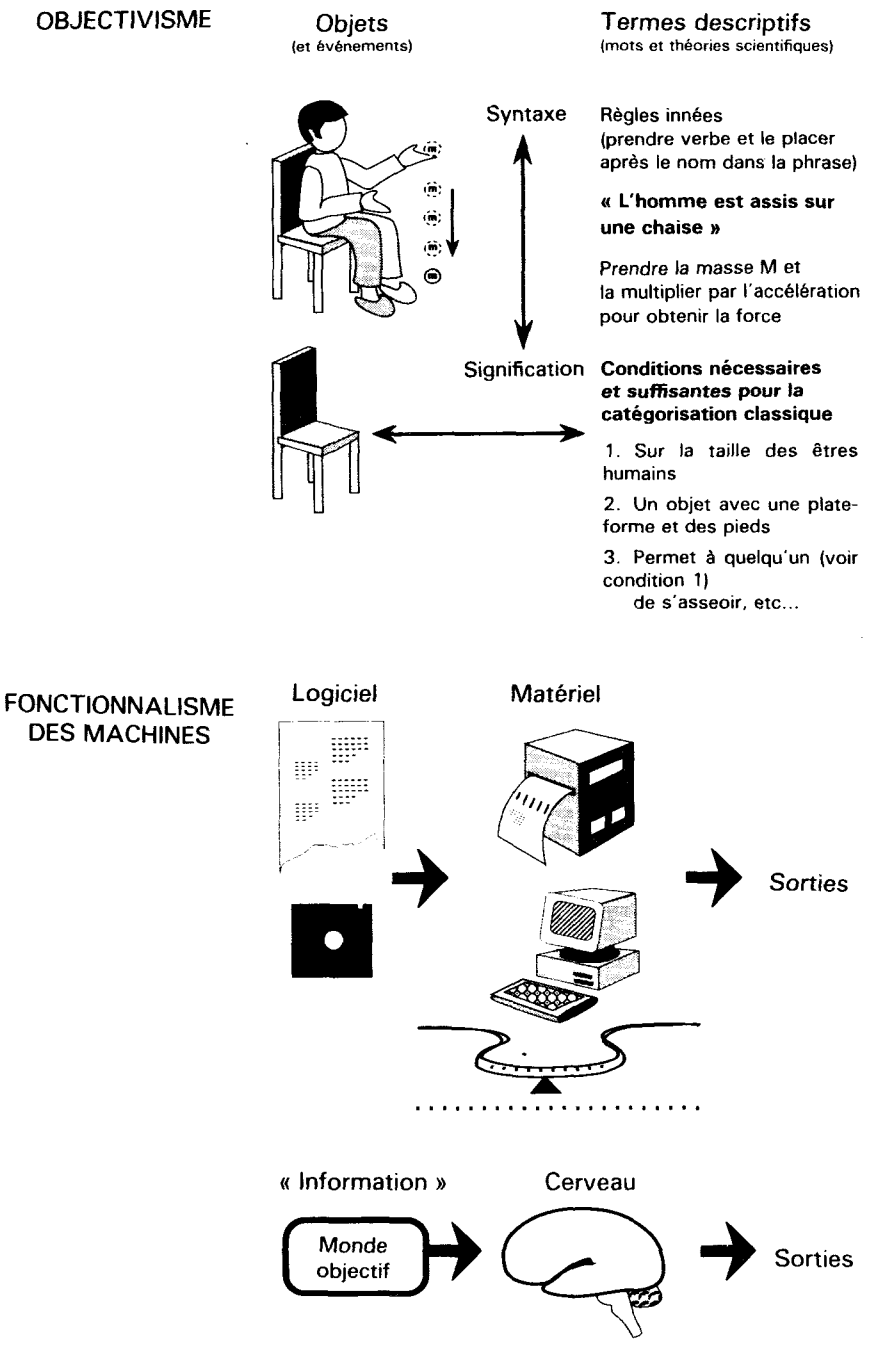


FIGURE P-5

Quelques aspects de l'objectivisme et du fonctionnalisme.

paraître à première vue, il est malheureusement incohérent et en désaccord avec les faits. Pourquoi donc est-il apparu ? Parce que, dans les sciences dures, il permet d'aller très loin. En effet, en ce qui concerne une grande partie de la chimie et de la physique classiques, avoir placé l'esprit hors de la nature constitue *effectivement* une sage précaution. Et en physique, le fait qu'un grand nombre de progrès majeurs aient été accomplis est largement dû à l'utilisation du raisonnement formel rigoureux qui se trouve au cœur des mathématiques et de la logique.

Vers la fin du XIXe siècle et le début du XXe, les recherches approfondies en logique mathématique menées par Gottlob Frege, Giuseppe Peano, Alfred North Whitehead et Bertrand Russell, suivies par les travaux de Stephen Kleene, Emil Post, Alonzo Church, Alan Turing et Kurt Gödel, constituèrent de véritables exploits dans le domaine de l'analyse par l'être humain de la « mécanique » du raisonnement logique. Pendant mes études universitaires, je fus charmé par l'élégance de tout cela. Je passai de longues soirées en compagnie de ces formidables concis de hiéroglyphes logiques à la couverture bleu foncé que sont les *Principia Mathematica* de Whitehead et Russell, convaincu que j'étais, par leur sécheresse même, de me trouver parmi les heureux élus. Dommage qu'à l'époque, je n'aie eu personne pour me parler du côté humain de ces auteurs. Depuis, j'ai entendu dire qu'au cours de la rédaction de ces volumes, Whitehead – qui habituellement était un homme paisible – dit un jour à son collègue, au tempérament plus bagarreur : « Bertie, le monde est fait de simples d'esprit et d'esprits confus, et je te laisse le soin de décider de quel côté tu te trouves. » Le mathématicien G. C. Rota a récemment critiqué de façon cinglante la propension excessive de certains philosophes à s'appuyer sur le formalisme et l'axiomatique, et à singer la clarté des mathématiques en adoptant un mode symbolique de discussion (voir les références à ce travail dans la bibliographie correspondant au chapitre 14).

Par la suite, l'avènement des ordinateurs, qui fut en partie dû à ces recherches, renforça les idées d'efficacité et de rigueur, ainsi que l'engouement pour le raisonnement logique qui caractérisait déjà une grande partie de la physique. Le formalisme logique « propre et net » qui se trouvait derrière les ordinateurs, le lien avec la physique mathématique et le succès des sciences dures semblaient extensibles à l'infini. Il y avait une tendance naturelle à arrêter l'analyse philosophique de l'exploration scientifique à la surface du corps humain (la peau et ses récepteurs). On pouvait analyser les

comportements, mais pas l'expérience phénoménale. Ainsi, la science pouvait demeurer « extensionnelle », comme le disait W. V. Quine, et on pouvait déclarer comme lui que « être, c'est être l'une des valeurs d'une variable ».

La conception computationnelle ou représentationnelle considère la nature comme vue de l'extérieur. Elle est imposante et *semble* permettre d'établir une charmante correspondance entre l'esprit et la nature. Cependant, une telle correspondance n'est charmante que tant qu'on laisse de côté la question de savoir comment se révèle concrètement l'esprit chez des êtres humains pourvus d'un corps. De fait, lorsqu'on l'applique à l'esprit *in situ*, cette conception devient intenable.

Les problèmes posés par la conception computationnelle de l'esprit, fondée sur les représentations mentales, sont multiples, mais ils peuvent être regroupés en huit types principaux (tableau P-1). Une telle classification n'est pas seulement commode ; elle fournit en outre un plan de bataille pour qui veut s'attaquer à cette conception de l'esprit. Je conseille aux lecteurs intéressés de consulter la bibliographie, car ils y trouveront des références leur permettant d'approfondir leur compréhension des travaux des auteurs cités dans le tableau P-1. Quant à moi, comme je pars de l'hypothèse que les lecteurs suivront mes conseils, je n'aborderai ici que brièvement les questions énumérées dans ce tableau. Mon but est d'ébaucher les arguments critiques majeurs contre le fonctionnalisme et l'objectivisme, et non d'en donner une description exhaustive.

Les catégories : un obstacle majeur aux visions fonctionnalistes de la cognition

L'un des plus grands défis auxquels est confrontée la conception fonctionnaliste des représentations mentales provient des travaux de philosophie et de psychologie sur la façon dont nous classons les choses par catégories. La majeure partie de ces travaux concerne la catégorisation conceptuelle chez l'être humain, mais quelques-uns traitent également de la catégorisation perceptive à la fois chez l'homme et l'animal. La conclusion générale la plus frappante qui se détache des diverses analyses et études est que les êtres humains ne classent pas les choses et les événements d'après des catégories classiques. Je rappelle que les catégories classiques sont celles dans

lesquelles l'appartenance est définie d'après des conditions nécessaires et suffisantes (figure P-5).

TABLEAU P-1

Quelques problèmes posés par la notion de représentation mentale *

1. La perception et la raison ne sont pas régies par des catégories classiques. La biologie (notamment les travaux de Darwin) montre que l'essentialisme est faux (Rosch, Wittgenstein, Lakoff, Mayr). Similitude et catégorisation ne sont pas une seule et même chose.
2. La pensée n'est pas transcendante ; elle dépend du corps et du cerveau. Elle est *incarnée*. La signification se crée par rapport aux besoins et aux fonctions corporelles. L'esprit n'est pas un miroir de la nature (Putnam, Millikan, Langacker, Lakoff, Johnson, Searle, Edelman).
3. La mémoire ne peut être décrite par des codes internes ou des systèmes syntaxiques. De plus, on a besoin d'un moi et d'une conscience d'ordre supérieur pour rendre pleinement compte de ses manifestations linguistiques (Searle, Shanon, Gauld, Edelman).
4. On acquiert le langage en interagissant avec d'autres au cours de phénomènes d'apprentissage qui déclenchent la formation de liens entre les significations et la phonétique. Pour cela, il faut que des systèmes conceptuels et des valeurs soient déjà en place (Pinker, Johnson, Edelman).
5. Chaque esprit crée sa propre version de la réalité à travers des interactions sociales et linguistiques, et la réalité, tout comme la biologie elle-même, dépend d'événements historiques (Searle, Putnam).
6. Non seulement les calculs sont désincarnés, mais ils ne permettent pas à eux seuls d'établir des relations significatives entre des symboles et des entités du monde réel (Searle).
7. La cognition puise son contenu dans l'identification de fonctions appropriées au sein d'un système qui dépend de l'histoire évolutive. Chaque partie d'une fonction appropriée possède une explication « normale », qui précise comment ce système est parvenu, au cours de l'évolution, à assurer cette fonction. Le « rationalisme du sens » – l'attribution d'une signification venue d'en haut – n'est pas tenable (Millikan).
8. La structure, la fonction et la variabilité du système nerveux, ainsi que son évolution et son développement, sont incompatibles avec la vision fonctionnaliste (Edelman).

* *Les noms entre parenthèses sont ceux des auteurs dont les travaux figurent dans la liste bibliographique à la fin de cet ouvrage. Prière de se reporter aux travaux cités dans cette liste pour des explications plus détaillées.*

Wittgenstein a été l'un des premiers à avoir une pensée critique à ce sujet. En réfléchissant aux ressemblances familiales, il remarqua que les membres d'une catégorie peuvent être apparentés les uns aux autres même si certains d'entre eux n'ont en commun aucune des propriétés qui définissent classiquement leur catégorie (figure P-6, à droite). (Supposez que n propriétés soient distribuées

parmi les membres de l'ensemble et que *m* propriétés suffisent pour être membre, avec *n* supérieur à *m*. Si *m* des *n* propriétés assurent l'appartenance, deux membres pourront n'avoir aucune propriété commune. Cela définit en partie ce qu'on appelle un ensemble polymorphe.) Wittgenstein se pencha également sur plusieurs autres idées fascinantes – sur le fait que certaines catégories puissent présenter des *degrés* d'appartenance tout en étant dépourvues de frontières nettes, et que certaines autres puissent avoir des membres plus centraux ou plus « prototypiques » que d'autres.

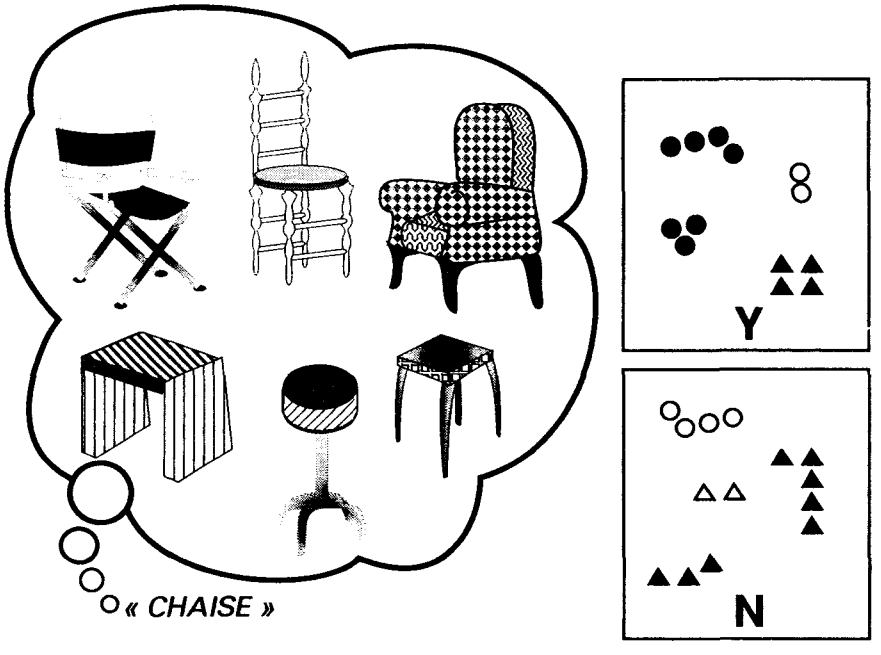


FIGURE P-6

Catégorisation et ensembles polymorphes. À gauche *: les chaises ne sont pas nécessairement caractérisées par des critères nécessaires et suffisants (catégories classiques).* À droite *: règle polymorphe d'appartenance à un ensemble lorsque la catégorisation classique ne s'applique pas. Dans cet exemple, les membres de l'ensemble (groupe indiqué par un* Y*, signifiant « yes ») possèdent deux des trois propriétés suivantes : rondeur, couleur pleine et symétrie bilatérale. Ceux qui n'y appartiennent pas (groupe indiqué par un* N*) n'ont qu'une seule de ces propriétés. La figure provient des expériences effectuées par Ian Dennis et ses collaborateurs.*

Depuis l'époque de Wittgenstein, un certain nombre d'études effectuées par des psychologues ont confirmé ses idées. Les plus remarquables de ces travaux sont ceux de Brent Berlin et Paul Kay, qui ont montré que les catégories humaines de couleurs présentent une « centralité » et des degrés d'appartenance ; ceux de Roger Brown,

qui a montré que les enfants commencent par nommer les choses à un niveau qui n'est ni le plus général ni le plus spécifique ; et ceux d'Eleanor Rosch et de ses collègues, dont les travaux sont peut-être les plus généraux, et qui ont transformé l'analyse des catégorisations en un outil de recherche aux multiples applications.

Les travaux de Rosch ont démontré l'existence de ressemblances familiales, de centralité et de prototypisme. Ainsi, des catégories telles que « rouge » ont des frontières floues, mais elles contiennent néanmoins des membres centraux dont le degré d'appartenance, s'il était mesuré par rapport à une échelle allant de zéro à un, serait égal à un. Ces catégories sont dites graduées. En revanche, des catégories telles que « oiseau » ont des frontières très nettes, mais à l'intérieur de ces frontières, certains oiseaux sont considérés comme étant plus représentatifs que d'autres – plus « prototypiques ». L'identification des membres d'une catégorie est souvent structurée autour d'un niveau de base – un niveau qui, chez les sujets testés par Rosch, est mis en évidence par la facilité d'imaginer et de se rappeler l'appartenance, les actions et l'utilisation. Dans ce sens, « cheval » est une catégorie de niveau de base, mais pas « quadrupède ».

Si nous admettons l'existence de ressemblances familiales, nous ne serons pas surpris de constater que souvent il n'existe pas de relations hiérarchiques précises entre catégories supra-ordonnées et subordonnées. Le fait que les catégories soient d'origine hétérogène est compatible avec cela : les propriétés que les humains utilisent effectivement pour déterminer l'appartenance à des catégories sont interactives et dépendent de diverses variables biologiques, culturelles et environnementales.

Ce travail empirique a été réalisé sur des sujets humains. Et, bien que certains de ses aspects aient été contestés de temps à autre, il a été, en général, confirmé. Plus récemment, Lance Rips a montré que ni la ressemblance ni la « typicalité » ne rendent totalement compte du *degré* d'appartenance à une catégorie, et que le raisonnement en jeu dans la détermination de l'appartenance ne relève souvent pas de la logique. Lawrence Barsalou a en outre montré que les catégories particulières ne sont même pas représentées par des concepts invariants. Il existe une énorme variabilité au niveau des concepts qui représentent une catégorie donnée : des individus différents ne la représentent pas de la même manière, et le même individu modifie sa conception de l'appartenance à cette catégorie suivant les contextes. D'ailleurs, les études fondatrices de Daniel

Kahneman et Amos Tversky sont compatibles avec ces idées ; elles ont permis de montrer que les décisions prises par des sujets humains et leurs jugements concernant l'appartenance à des catégories violent souvent des règles de probabilité telles que la règle de conjonction, qui veut qu'une conjonction ne soit jamais plus probable que l'un ou l'autre de ses composants. Dans certains contextes, il arrive en effet que des sujets soient sûrs que la conjonction est plus probable.

Je ne me suis intéressé ici qu'aux catégories conceptuelles, car j'ai déjà abordé les catégories perceptives dans le corps de ce livre. Mais ce que nous venons de voir suffit pour affirmer que si ces travaux sont exacts, le modèle objectiviste de la relation entre l'esprit et le monde est dans de bien mauvais draps. Par exemple, si les catégories qui présentent une centralité et un prototypisme, telles celles des couleurs, existent en plus des catégories classiques, alors la vision objectiviste n'est pas la bonne. Pis encore, le modèle objectiviste ne permet pas d'expliquer le fait que *certains symboles ne correspondent à aucune catégorie existant dans le monde.* Ainsi, certains travaux de psychologie indiquent, par exemple, que lorsqu'on assimile l'esprit à un ordinateur, on n'arrive pas à rendre compte des catégories de l'esprit et du langage (cf. n'importe quel poème) qui ne reflètent aucune catégorie existant dans le monde réel. Par ailleurs, les individus peuvent comprendre les événements et les catégories de diverses manières, qui sont parfois incohérentes. Comme l'a fait remarquer Mark Johnson, la métaphore et la métonymie sont des modes de pensée majeurs. La métaphore consiste à transférer les propriétés d'une chose sur une autre chose, issue d'un domaine différent ; la métonymie permet à une partie ou à un aspect d'une chose de représenter la chose dans son ensemble. Toutes deux sont incompatibles avec la conception objectiviste.

Tout cela pose des problèmes à la notion de représentation mentale. En effet, pour fonctionner, le mentalais requiert l'existence d'une liaison précise et dépourvue d'ambiguïté avec le monde extérieur. Mais, souvent, il est impossible d'établir la signification de cette manière, et une telle liaison ne peut donc pas exister. Les objets du monde ne sont *pas* étiquetés à l'aide de dimensions ou de codes, et la façon dont les individus en effectuent la partition varie d'un individu à un autre et d'un instant à un autre. En fait, la sémantique figée des représentations mentales ne peut rendre compte du fait que le monde produit des nouveautés et, comme nous le verrons lorsque j'aborderai le langage, les codes bien définis ne peuvent épuiser le sens des expressions linguistiques. La signi-

fication refuse tout simplement de se laisser ligoter par un ensemble fixe de termes, écrits en utilisant un système de codage spécifique. Alors que les représentations doivent demeurer figées, les comportements changent en fonction du contexte (de façon inexplicable, pour qui les voit sous l'angle objectiviste).

Si ce que je dis est exact, cela signifie que l'esprit n'est pas un miroir de la nature. La pensée ne se limite pas à la manipulation de symboles abstraits dont la signification est justifiée par référence univoque à des choses du monde réel. Les catégories classiques ne servent à rien dans la plupart des cas de catégorisation conceptuelle, et elles ne parviennent pas à rendre correctement compte de la façon dont les êtres humains effectuent des catégorisations dans la pratique. Il n'existe pas de correspondance univoque entre le monde et notre catégorisation de ce monde. Autrement dit, l'objectivisme ne marche pas.

La mémoire et le langage

La mémoire et ses rapports avec le moi et le langage constituent une autre source d'embarras pour la conception computationnelle, ou fonctionnaliste, de l'esprit. Au cours de la section suivante, j'aborderai certains aspects particuliers du langage, mais pour l'instant, il vous suffira de remarquer que les mots du langage naturel ne ressemblent pas aux termes d'un langage informatique. J'ai déjà attiré l'attention, au cours de la section précédente, sur le fait que tout calcul est syntaxique par nature et que, par conséquent, contrairement à l'utilisation des mots au sein d'une communauté linguistique, il ne peut avoir de sens qu'en présence d'un programmeur. De plus, les fonctionnalistes parlent souvent d'attitudes propositionnelles – croyances, désirs, souhaits. Mais, comme l'a fait remarquer Putnam, les croyances et les désirs ne peuvent être individualisés qu'en présence de références à un environnement non figé – à un environnement qui n'a pas été défini d'avance.

A cela s'ajoute encore un autre problème : la mémoire humaine ne ressemble en rien à une mémoire d'ordinateur. Comme nous l'avons déjà vu, les codes internes et les systèmes syntaxiques sont incapables de décrire la mémoire humaine de façon satisfaisante. Selon les cas, on a dit de la mémoire qu'elle était épisodique (relative à des événements passés de la vie de l'individu), sémantique (relative au langage), procédurale (relative aux actes moteurs), déclarative

(se référant aux énoncés), etc., ce qui n'a pas manqué, parfois, de semer la confusion. *Mais la mémoire est une propriété systémique* : elle varie selon la structure du système dans lequel elle s'exprime. De plus, dans les systèmes biologiques, il ne faut pas confondre la mémoire avec les mécanismes nécessaires à son établissement, telles les modifications synaptiques. Et surtout, la mémoire biologique n'est pas une copie conforme, une trace qui a été codée pour représenter son objet.

Quelle que soit sa forme, la mémoire humaine fait appel à un ensemble de liaisons apparemment non figées entre des sujets et un riche tissu de connaissances antérieures qui ne peuvent être représentées proprement par le langage appauvri de l'informatique – par des expressions telles que « stockage », « recherche d'information », « entrée », « sortie ». Pour avoir une mémoire, on doit être capable de reproduire des résultats ou des comportements passés, d'affirmer des choses, de relier des thèmes et des catégories à sa propre position dans le temps et dans l'espace. Et pour ce faire, on doit posséder un moi, et qui plus est un moi conscient. Autrement, nous serions obligés de postuler l'existence d'un petit bonhomme chargé d'effectuer les recherches d'information dans la mémoire (en informatique, nous, programmeurs, sommes les petits bonshommes). Comment utiliser le modèle fonctionnaliste d'un esprit algorithmique autrement qu'en faisant appel à une infinité d'homoncules, emboîtés les uns dans les autres ?

À travers la question de l'homoncule, nous sommes amenés à considérer l'un des grands problèmes qui se posent concernant la question de l'esprit : comment rendre compte de l'intentionnalité elle-même ? Nous avons déjà montré qu'une construction sémantique formelle, dénuée d'ambiguïté, ne pourra jamais attester de l'état réel des choses. Pourtant, un grand nombre d'aspects causatifs de nos états mentaux dépendent du contenu sémantique. En fait, comme l'a fait remarquer Searle, le contenu sémantique n'a aucun sens en l'absence d'intentionnalité, c'est-à-dire de la capacité de faire référence à d'autres états ou objets. Et pour qu'on en soit capable, il faut que les représentations formelles deviennent des représentations intentionnelles. Chez les êtres humains, cela exige la présence d'une conscience et d'un moi – d'une sensibilité individuelle biologiquement fondée, d'une première personne. Aucune théorie de l'esprit digne de ce nom ne peut échapper à cette question, qui n'est pas seulement une question de langage mais aussi un grand problème de biologie. Continuons donc implacablement notre

quête, en nous tournant enfin vers certaines des questions biologiques qui ne peuvent être réconciliées avec la vision fonctionnaliste de l'esprit.

Les leçons de la *biologie*

Les efforts déployés par Darwin pour comprendre l'origine des espèces permirent l'avènement d'une grande révolution de la pensée. Avec sa théorie de la sélection naturelle, Darwin fut le premier à penser en termes de populations. Cela revient à considérer, selon les mots d'Ernst Mayr, que la variance au sein d'une population est réelle, c'est-à-dire *qu'il ne s'agit pas d'une erreur* (voir la figure 5-2). La sélection naturelle agit sur la variabilité entre individus d'une même population. Comme l'a montré Mayr, l'apparition d'une espèce résulte souvent de la présence de barrières sexuelles et géographiques à la propagation des variantes, et peut même être accidentelle.

Le concept d'espèce qui résulte de ce mode de pensée est au centre de toutes les idées sur la catégorisation. En effet, les espèces ne sont pas des « types naturels » ; leur définition est relative, elles ne sont pas homogènes, aucune condition préalable n'est nécessaire à leur établissement et elles n'ont pas de frontières bien définies.

Ainsi, le fait de penser en termes de populations a porté un coup mortel au raisonnement typologique ou essentialiste, c'est-à-dire à l'idée selon laquelle l'« essence » des espèces existe avant les organismes particuliers, ou exemplaires, de ces espèces. L'essentialisme, dont la formulation la plus claire est due à Platon, et qui s'est trouvé reflété dans la plupart des philosophies idéalistes depuis lors, est intimement lié à la notion de catégories classiques. Mais la biologie nous montre que, même s'il est possible d'établir une taxonomie des créatures vivantes, l'essentialisme est une conception fausse. Et, vu mes précédentes remarques, il est probablement tout aussi faux en ce qui concerne ses vues sur l'esprit.

Searle, Lakoff, Johnson et d'autres (dont moi-même) ont fait remarquer que la pensée n'est pas transcendante, mais qu'elle dépend au contraire intrinsèquement du corps et du cerveau. Ce point de vue est diamétralement opposé à celui du fonctionnalisme, qui suppose que la mise en œuvre du logiciel est *indépendante* du matériel. En effet, d'après ceux qui rejettent le fonctionnalisme, l'esprit est incarné. Certains diktats corporels doivent donc néces-

sairement être suivis par l'esprit. La perception gestaltiste est certainement l'un de ces diktats, puisque les catégories d'une figure gestaltiste (voir, par exemple, la figure 4-2), qui ne sont pourtant validées par aucune configuration *issue du monde réel*, sont souvent impossibles à rectifier. Les figures gestaltistes, les images mentales, les mouvements corporels et l'organisation des connaissances doivent tous, dans une certaine mesure, être le résultat des contraintes imposées par l'évolution et le développement.

La syntaxe et la sémantique des langues naturelles ne sont pas simplement des cas particuliers de syntaxe et de sémantique formelles, dont les modèles ont une structure mais sont dépourvus de signification. Selon le point de vue biologique, l'attribution de signification aux symboles ne s'effectue *pas* de façon formelle ; au contraire, on suppose que les structures symboliques ont un sens *dès le départ*. Cela est dû au fait que les catégories sont déterminées par la structure corporelle et par l'utilisation adaptative à l'issue de l'évolution et des comportements. Les symboles cognitifs doivent correspondre à l'appareil conceptuel contenu dans les cerveaux réels. Les bases de la vérité et de la connaissance proviennent de cet appareil et trouvent leurs racines les plus anciennes dans les systèmes de valeurs mis en place au cours de l'évolution. Selon les défenseurs de ce point de vue, dont Lakoff, Johnson, Modell et moi-même, lorsque les symboles ne correspondent pas directement à ce qui existe dans le monde, les êtres humains utilisent des métaphores et des métonymies pour établir des liens, hormis l'imagerie et la perception des schémas corporels.

À travers les interactions culturelles et linguistiques, l'esprit *crée* certains aspects de la réalité. Tout comme la biologie elle-même, ces interactions dépendent des événements historiques. J'aborderai ces questions au cours de la section suivante, lorsque je parlerai du langage et de son acquisition.

Hormis l'incarnation, il reste encore une question clé : celle de la fonction. Millikan, qui a réfléchi de façon très approfondie à cette question, a exprimé ses réflexions de façon très incisive en ce qui concerne l'esprit, le langage et « autres objets biologiques », selon ses propres mots. Les objets biologiques soumis à l'évolution ont des propriétés fonctionnelles différentes de celles, disons, des molécules. On ne parle pas de la fonction « anormale » d'une molécule en tant qu'objet chimique. En revanche, les objets biologiques ont une fonction appropriée, qui dépend de leur histoire évolutive. La fonction appropriée du cœur consiste à pomper du sang. De

même, il y a ce que Millikan appelle une explication « normale » à la production d'un tel objet au sein d'une espèce, et c'est cela qui rend compte de la ressemblance qui existe entre cet organe et les cœurs « normaux » pour cette espèce. Les cœurs peuvent fonctionner bien ou mal, et ceux qui fonctionnent mal sont anormaux. En revanche, les composés organiques font ce qu'ils font et, *quoi qu'ils fassent*, cela fait partie de leur « fonctionnement ».

Au cours de l'évolution, les fonctions qui rendent compte de la prolifération de survivants sont des fonctions appropriées, et elles sont associées à des explications « normales » qui rendent compte de la façon dont les survivants sont *historiquement* parvenus à assurer cette fonction. Et le plus drôle, c'est que des états et des activités peuvent avoir des fonctions appropriées sans pour autant les assurer et qu'ils peuvent même avoir des fonctions appropriées sans pour autant contribuer à d'autres fonctions appropriées, comme le voudrait leur explication « normale ». Cela est dû au fait que, dans les systèmes sélectifs, les phénomènes historiques aboutissent soit à des échecs, soit à des succès inespérés.

Millikan considère la psychologie comme une branche de la biologie, et je pense qu'elle a raison. Elle affirme que la cognition puise son contenu dans l'identification des fonctions appropriées, ce qui constitue une affirmation de taille. Pour elle, chaque ensemble de fonctions possède une explication « normale » qui décrit comment le système parvient à remplir cette fonction. Son point de vue permet de situer la cognition dans un contexte physiologique (par exemple, celui des systèmes de valeurs que j'ai évoqués au cours de ce livre) sans pour autant lui interdire de servir de base à une théorie des croyances et des désirs. Et en fait, contrairement aux attitudes propositionnelles du fonctionnaliste, une telle théorie de l'intentionnalité ne diffère pas foncièrement des pratiques et des références de la psychologie populaire habituelle (la façon dont nous caractérisons habituellement la fonction mentale dans la vie quotidienne). Millikan considère le cerveau comme un manipulateur de symboles et un moteur sémantique. Cela est dû au fait que, conformément à son point de vue, les croyances et les désirs sont « normalement » manipulés d'après les différences significatives (c'est-à-dire *corporellement* significatives) existant entre eux et aussi d'après les différences existant entre leurs fonctions appropriées. Selon son analyse, l'évaluation de la signification et de la vérité provient de cette voie, et non d'une attribution sémantique à travers des correspondances qui sont établies par les « rationalistes du sens »

(c'est ainsi qu'elle appelle ceux qui défendent le point de vue opposé au sien).

D'après tous les arguments que je viens de décrire, il s'ensuit que les données de la biologie nous obligent à conclure que l'esprit n'est pas transcendant. Il ne peut donc pas regarder le monde de l'extérieur. L'essentialisme n'est pas une position tenable, pas plus que ne le sont le fonctionnalisme, l'objectivisme ou la forme de « réalisme computationnel » qui considère que l'esprit est une machine. De plus, il existe une autre source profonde de problèmes, que j'ai décrite dans les premiers chapitres de ce livre et ailleurs : la variabilité interne des structures et des fonctions du système nerveux, d'une part, et la façon dont le cerveau développe sa connectivité anatomique, d'autre part – en se fondant sur des corrélations avec des événements survenant dans le monde –, sont toutes deux incompatibles avec le point de vue fonctionnaliste.

Les cercles vicieux gravés dans le paysage cognitif peuvent être rompus par les données qui sous-tendent l'analyse qui vient d'être faite. Mais il ne suffit pas de dire que l'esprit est incarné pour rendre compte de la signification et de la mémoire. La question est d'expliquer comment les choses se passent. Et ensuite, de voir comment cette explication peut rendre compte du développement du moi et de la conscience. C'était là la tâche que j'avais entreprise dans le corps de ce livre et, pour m'en acquitter, j'avais dû considérer le langage par rapport à la conscience d'ordre supérieur. Cependant, certains autres aspects techniques concernant spécifiquement le langage doivent être abordés dans le contexte des arguments exposés dans cette postface. C'est ce que nous allons faire maintenant.

LE LANGAGE : POURQUOI L'APPROCHE FORMELLE ÉCHOUE

Tout d'abord, je voudrais montrer en quoi les conceptions formelles du langage sont incompatibles avec ce que j'ai déjà dit sur les catégories. Ensuite, je voudrais évoquer quelques-uns des modèles cognitifs et des grammaires proposés qui s'accordent mieux avec ce que l'on sait sur la catégorisation. Mon but est de comparer ces deux points de vue – le formel et le cognitif – pour tenter de donner au lecteur un aperçu des grandes différences existant au niveau de leurs hypothèses de départ.

L'étude du langage pose un énorme défi, et le domaine de la

linguistique proprement dite est extrêmement complexe. Je ne tenterai pas ici de détailler ce type de travaux, car je ne les connais pas suffisamment à fond. Mais heureusement, pour ce qui nous occupe ici, un petit nombre d'indications suffiront. Je les décrirai brièvement pour aborder ensuite mon propos principal : le fait que les approches formelles de la grammaire tombent sous le même couperet que les approches objectiviste et strictement fonctionnaliste de la psychologie.

Connaître un langage, c'est être capable de produire des sons ou des gestes porteurs de sens, et de les comprendre lorsqu'ils sont produits par d'autres. En général, la relation entre la forme et la signification au sein d'un langage est arbitraire. L'une des caractéristiques frappantes du langage est la créativité qu'il autorise : tout individu versé dans un langage donné est capable de former et de comprendre des expressions et des phrases complètement inédites. Tout aussi frappant est le fait que, la plupart du temps, les individus peuvent faire la différence entre locutions propres et impropres du point de vue grammatical.

Les linguistes tentent de construire des théories de la grammaire des locuteurs. En ce sens, être grammatical signifie se conformer à des règles descriptives issues de l'usage courant du langage. Dans son sens le plus large, cependant, la grammaire inclut l'étude des lois phonétiques (les lois du système de sons), morphologiques (dans ce contexte, les lois de formation des mots) et sémantiques (les lois du système de significations). L'ensemble de ces lois est appelé « grammaire universelle », expression inventée par Noam Chomsky et aujourd'hui consacrée. D'après les idées fondatrices de Chomsky, toutes les langues possèdent un ensemble de propriétés grammaticales communes, qui constituent cette grammaire universelle.

Chomsky a également suggéré que, dans la mesure où seuls les êtres humains ont un langage et que la performance réelle des enfants est sous-déterminée par leurs compétences testables, il doit exister chez eux un « dispositif d'acquisition du langage » inné. Il est important de remarquer ici que la parole est une compétence *acquise*, qui se développe du fait de l'appartenance à une communauté linguistique. Il faut effectuer beaucoup de catégorisations pour réussir à parler. Des concepts et des intentions doivent se développer, des expressions doivent être formulées conformément à la grammaire et à la phonétique, et il faut être capable d'articuler, de comprendre et de garder la trace de ses propres paroles au cours des échanges avec autrui.

Pour ce faire, en tant qu'interlocuteur, il faut un principe coopératif du type de celui qui a été décrit par H. P. Grice. Nous devons être informatifs précisément au niveau requis, et pas plus ; nous devons être brefs, méthodiques et non ambigus. Nous devons être réceptifs aux indices nous permettant de savoir quand arrive notre tour de parler. De plus, nous devons définir un « ici et maintenant » ou un « là-bas et à ce moment-là ». Une telle déictique – c'est ainsi qu'on l'appelle – permet de situer à la fois les interlocuteurs et les objets dans l'espace. Enfin, ce que nous avons l'*intention* de communiquer doit également s'exprimer dans des actes de paroles adéquats. Le fait de parler est généralement une question de tact autant que de tactique.

L'acquisition du langage au sein d'une communauté linguistique ne s'effectue pas de la même manière chez les enfants et chez les adultes. D'ailleurs, acquérir un langage et l'utiliser ne sont pas nécessairement la même chose. Pour le voir, on doit faire intervenir à la fois l'étude des connaissances linguistiques, ou psycholinguistique, et l'étude des bases biologiques et neuronales du langage, ou neurolinguistique. Nous avons analysé les différences entre l'acquisition et l'utilisation du langage au chapitre 12. Je vous le rappelle ici afin de vous éviter de confondre l'acquisition d'une langue avec l'utilisation de cette langue permise par un entraînement.

Tout cela me permet d'introduire le problème du rapport entre pensée et langage. Nous devons esquisser le tableau de la relation existant entre systèmes de concepts et langage de façon claire. La maîtrise du langage dépend-elle de l'existence d'un riche système incarné de concepts ? Ou bien est-elle plus ou moins autonome, ce qui veut dire qu'elle se développe à travers un dispositif d'acquisition du langage ?

L'une des approches les plus répandues et influentes de ces questions fondamentales est due aux travaux fondateurs de Chomsky. Dans son approche, fondée sur les systèmes formels, la principale hypothèse consiste à dire que les règles de syntaxe sont indépendantes du système sémantique. Vu sous cet angle, le langage est indépendant du reste de la cognition. Je me permets de ne pas partager cette idée.

L'ensemble de règles formulées en partant de l'idée que la grammaire est un système formel est essentiellement algorithmique. Dans un tel système, le sens n'intervient pas. La « grammaire générative » de Chomsky (figure P-7) part du principe que la syntaxe est indépendante de la signification et que les facultés linguistiques

sont indépendantes des facultés cognitives externes. Il est donc vain, en particulier, de tenter d'infirmer cette définition de la grammaire en faisant référence à des données sur la cognition en général. Lorsqu'un langage est défini comme un ensemble de chaînes de symboles non interprétés engendrés par des règles de production, il ressemble à un langage d'ordinateur ; et, pour rendre les symboles sémantiquement significatifs, il faut les mettre en correspondance avec le monde réel ou avec un langage de la pensée ou mentalais.

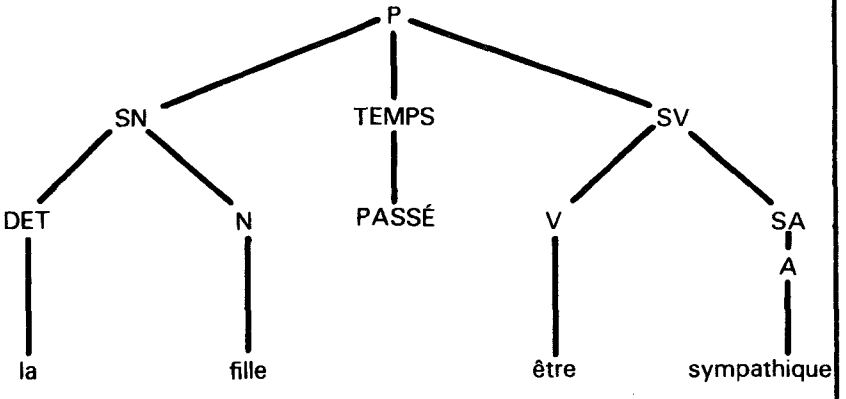
Forts de ce qui précède, nous pouvons à présent conclure que la conception sous-jacente à ce point de vue n'est rien d'autre que la conception objectiviste : les catégories sont classiques et les significations sont engendrées par attribution non ambiguë de symboles à des entités du monde réel. Cela revient à définir le langage et la grammaire. Et, d'après cette définition, le langage se heurte à tous les obstacles rencontrés par la conception objectiviste. Le problème n'est pas simplement dû au fait que cette conception soit en désaccord avec les données empiriques concernant la catégorisation. Il est également dû au fait que l'objectivisme omet de préciser que le langage sert à transmettre les pensées et les sentiments d'individus qui pensent déjà, indépendamment du langage.

Le dispositif d'acquisition du langage fut proposé par Chomsky afin de répondre à la question de savoir comment il se fait qu'un enfant, qui est apparemment incapable de comprendre nombre de choses très simples, puisse maîtriser les complexités du langage. Mais un certain nombre d'observations semblent incompatibles avec la position chomskienne. Elles concernent la pensée et l'acquisition du langage chez l'enfant, telles qu'elles ont été décrites, par exemple, par Margaret Donaldson dans son livre intitulé *Children's Minds*. Donaldson fait remarquer que Chomsky a attiré l'attention, dans son domaine, vers les études concernant la façon dont l'enfant acquiert une connaissance de la grammaire. Par conséquent, les linguistes ont recueilli et interprété les données sur ce que les enfants disent par rapport à un ensemble de règles qui aurait engendré ces énoncés. Mais ce faisant, ils ont mis de côté bien des choses – y compris, souvent, ce que l'enfant voulait effectivement dire et ce qu'il ou elle comprenait.

Comme le rappelle Donaldson, John MacNamara a suggéré que les enfants sont capables d'apprendre un langage *parce* qu'ils ont commencé par comprendre des situations faisant intervenir des interactions humaines. Les enfants *commencent* par comprendre les choses et, surtout, ils comprennent ce que les gens font. La synthèse

GRAMMAIRE GÉNÉRATIVE

MARQUEURS SYNTAGMATIQUES : symbolisent l'analyse de la phrase « la fille était sympathique »



Symboles utilisés pour les marqueurs syntagmatiques

P	= phrase
SN	= syntagme nominal
DET	= déterminant
N	= nom
TEMPS	= marqueur de temps de verbe
SV	= syntagme verbal
V	= verbe
SA	= syntagme adjectif
A	= adjectif

FIGURE P-7

Arbre typique d'une grammaire générative, utilisé pour mettre en évidence et analyser la syntaxe. D'après Chomsky, la présence des règles d'une grammaire universelle est garantie, chez l'être humain, par un dispositif inné d'acquisition du langage, qui opère sur une telle syntaxe ou sur l'un de ses exemplaires modernes. Le lien avec la signification est assuré par l'hypothèse objectiviste (voir la figure P-5). Chomsky a remplacé cette analyse grammaticale par une théorie plus récente, dite du gouvernement et du liage, mais les hypothèses sous-jacentes demeurent inchangées.

de Donaldson met clairement en évidence le fait que les enfants sont capables de voir les choses en se mettant à la place de quelqu'un d'autre, et non pas seulement de leur propre point de vue. Ils ont une pensée logique et raisonnent par analogie dès l'âge de quatre ans environ, et avec beaucoup plus de compétence qu'on ne pensait auparavant. Il semble également que les enfants comprennent d'abord les situations et les intentions humaines, et seulement *après* ce qui se dit. Cela signifie que le langage n'est *pas* indépendant du reste de la cognition. Par conséquent, nous devons rendre compte de l'acquisition du langage non seulement du point de vue du développement, mais aussi du point de vue de l'évolution. J'ai déjà abordé longuement ce problème au chapitre 12, lorsque je me suis intéressé simultanément à l'incarnation des systèmes conceptuels *et* linguistiques.

Avant de me tourner vers des façons alternatives de considérer le langage, je voudrais évoquer une description presciente due au romancier Walker Percy, dont le recueil d'essais sur le langage a été publié dans un livre intitulé *The Message in the Bottle*. J'ai l'impression que le besoin de comprendre le langage et la signification furent au centre de sa vie et de son œuvre. Percy savait que la grammaire générative ou transformationnelle ne permettait pas d'expliquer le *langage*, et qu'il ne s'agissait là que de la description *formelle* d'une compétence ; autrement dit, aucune relation n'était requise entre cet ensemble d'algorithmes et ce qui se passe dans la tête d'un individu. Il avait également compris que la sensibilité individuelle est *symbolique* autant qu'intentionnelle. Ce que j'ai appelé la conscience d'ordre supérieur (voir le chapitre 12), est un « savoir avec » (con-science). Percy trouva à redire à la fois à l'approche behavioriste et à l'approche sémiotique du langage, qui ne prêtent aucune attention au caractère *intersubjectif* de tout acte linguistique. Il trouva également à redire à la philosophie phénoménologique, car elle « négligeait le deuxième larron » : il insista sur le fait que dans tout échange symbolique faisant intervenir des significations, on trouve une relation tétraédrique entre le symbole, l'objet et au moins deux êtres humains. Voici la façon dense et résonante dont Percy a lui-même exprimé cette idée : « L'acte de conscience consiste à faire intentionnellement en sorte qu'un objet soit ce qu'il est pour nous deux sous les auspices d'un symbole. » Percy décrit également le ravissement d'Hellen Keller lorsqu'elle apprit que l'eau était l'« eau » et son pressant désir de savoir ce

qu'« étaient » les autres choses. Le langage, comme le dit Percy, crée un monde, et non pas seulement un environnement.

Ce monde regorge d'intentionnalité, de projections, de sentiments, de préjugés et d'affection. Cela me rappelle une histoire à propos de deux touristes juifs, en visite pour la première fois en Israël. Après une journée, agréable mais épuisante, passée à Tel-Aviv, ils décident d'aller dans une boîte de nuit. Sur la scène, un comédien raconte des blagues en hébreu. Après en avoir entendu quelques-unes, l'un des touristes tombe de sa chaise, pris d'un fou rire. Son compagnon le regarde et lui demande : « Pourquoi ris-tu ? Tu ne comprends même pas l'hébreu. » Et l'homme qui se trouve par terre lui répond, en se tenant les côtes : « Parce que je fais confiance à ces gens. »

La sémantique formelle ne peut rendre compte d'une telle richesse. Mais alors, que pouvons-nous faire ? Une solution consiste à construire ce qu'on appelle une « grammaire cognitive », en partant des données cognitives au lieu de partir d'une analyse formelle. L'un des premiers pionniers de cette démarche fut Ronald Langacker, dont on peut consulter le livre intitulé *Foundations of Cognitive Grammar* pour connaître l'historique et les principes directeurs de son travail. Mais comme dans tous les sujets naissants, les terminologies varient. Et, plutôt que d'utiliser celle de Langacker, qui constitue dans un certain sens la terminologie « originale », je vais, pour des raisons de commodité, décrire et suivre les propositions de Lakoff – qui sont liées de près à celles de Langacker –, parce qu'elles fournissent des exemples plus proches de mes propres travaux en neurosciences. Prenons donc comme exemple la tentative de ce linguiste pour fournir un modèle de la cognition qui soit adapté aux données connues sur la catégorisation et à construire une sémantique fondée sur l'idée que la signification est incarnée.

Modèles cognitifs et sémantique cognitive : retour à la biologie

Lakoff a abordé le thème de la grammaire et de la sémantique d'une manière qui semble mieux correspondre aux données de la biologie et de la psychologie que les grammaires génératives. En partant de données concrètes sur la catégorisation, il en déduit que la signification résulte des mécanismes intrinsèques du corps et du cerveau. Il suggère que chaque être humain construit des modèles

cognitifs qui sont le reflet de concepts ayant trait aux interactions entre le corps-cerveau et l'environnement. C'est cette incarnation conceptuelle, affirme-t-il, qui aboutit à la formulation des catégories de niveau de base du type de celles décrites par Rosch.

Les modèles cognitifs sont *créés* par les êtres humains et, dans ce sens, ils sont idéalisés – c'est-à-dire abstraits. Mais ils dépendent de la formation d'images à l'issue d'expériences sensorielles, ainsi que de l'expérience kinesthésique – c'est-à-dire de la relation du corps avec l'espace. Lakoff suggère que l'exercice de ces fonctions aboutit à divers schémas d'images et de kinesthésie. Les schémas ont des propriétés qui se refléteront plus tard dans l'utilisation de la métaphore et de la métonymie. Je rappelle que la métaphore consiste à transférer, ou plaquer, une chose sur une autre, issue d'un domaine différent, tandis que la métonymie consiste à utiliser une partie ou un aspect d'une chose pour représenter la chose elle-même. Lakoff donne comme exemple de métaphore la phrase : « La colère est un animal dangereux. » Son exemple de métonymie est : « Le sandwich au jambon est parti sans payer. »

Ce qu'il importe de comprendre, c'est que les modèles cognitifs idéalisés font appel à l'incarnation conceptuelle et que celle-ci s'effectue grâce à des activités corporelles *antérieures au langage*. L'incarnation conceptuelle intervient dans la catégorisation et autorise l'hétérogénéité et la complexité des catégorisations humaines réelles. Ainsi, les catégories de l'esprit correspondent à des éléments des modèles cognitifs. Certains de ces modèles présentent différents degrés d'appartenance. D'autres incluent les catégories classiques et sont formés selon des conditions nécessaires et suffisantes (remarquez que cela n'est pas contradictoire, du moment que *tous* les modèles ne sont pas classiques !). Certains modèles sont métonymiques. Mais les modèles cognitifs les plus complexes correspondent à ce que Lakoff appelle des catégories radiales. Celles-ci comportent de nombreux modèles reliés à un centre. Bien que le fait de connaître la catégorie centrale ne permette pas de prédire les modèles (et les catégories) non centraux, ces derniers ont un rapport avec le centre ; Lakoff dit qu'ils sont « motivés » par le centre.

Ce type de propriétés autorise les degrés d'appartenance, les degrés de liaison avec le modèle central, les ressemblances familiales, les relations non hiérarchiques où les catégories de base dominent, et les effets de prototypisme. Ces derniers ne sont pas fondamentaux, mais proviennent de nombreuses sources – « scalaires », « classiques », « métonymiques » et « radiales ».

Sur cette toile de fond, Lakoff tente de bâtir une structure sémantique cognitive (figure P-8). Remarquez tout d'abord que la signification est déjà enracinée dans l'incarnation à travers les schémas d'images, les schémas kinesthésiques, les mots utilisés dans les métonymies et les relations catégorielles sous-jacentes aux métaphores. Mais cela ne suffit pas : le langage est censé être caractérisé par des modèles *symboliques*. Il s'agit là de modèles qui réalisent un appariement entre l'information linguistique et les modèles cognitifs, qui eux-mêmes constituent un système conceptuel *préexistant*. Dans la mesure où les modèles conceptuels préexistants sont déjà incarnés à travers leur liaison avec l'expérience corporelle et sociale, cette liaison n'est pas arbitraire. En revanche, l'attribution de ce type de liaisons à une grammaire générative en termes de représentations mentales *est* arbitraire : elle est faite d'en haut par le grammairien.

D'après cette conception sémantique cognitive, les catégories linguistiques présentent naturellement de fortes ressemblances structurelles avec leurs modèles cognitifs sous-jacents. Le langage utilise des mécanismes cognitifs généraux pour construire des modèles propositionnels, des modèles de schémas d'images, des modèles métaphoriques et des modèles métonymiques. Comme nous l'avons déjà vu, les modèles métaphoriques font correspondre une structure appartenant à un domaine à une structure issue d'un autre domaine. Cette mise en correspondance fait intervenir soit des schémas propositionnels, soit des schémas d'images. Les modèles métonymiques utilisent ces schémas et la métaphore pour définir une fonction entre un élément d'un modèle et un autre (par exemple, la relation entre une partie et un tout).

Dans *Women, Fire and Dangerous Things : What Categories Reveal about the Mind*, Lakoff utilise les travaux de son collègue Johnson (*The Body in the Mind : the Bodily Basis of Meaning, Imagination, and Reason*) pour construire une série de schémas, fondés sur des concepts incarnés, qui constituent la base du sens linguistique. Cet ensemble comporte notamment des schémas *récipients* (qui définissent une frontière, un « dedans et dehors »), un schéma *partie-tout*, un schéma de *liaison* (une chose reliée à une autre, comme par une ficelle), un schéma *centre-périphérie* (à l'instar de l'opposition entre le tronc et les extrémités), et un schéma *source-chemin-cible* (point de départ, chemin orienté, point situé à mi-chemin) comportant des schémas *haut-bas* et *avant-arrière*. Il procède ensuite à montrer que les métaphores sont motivées par la structuration

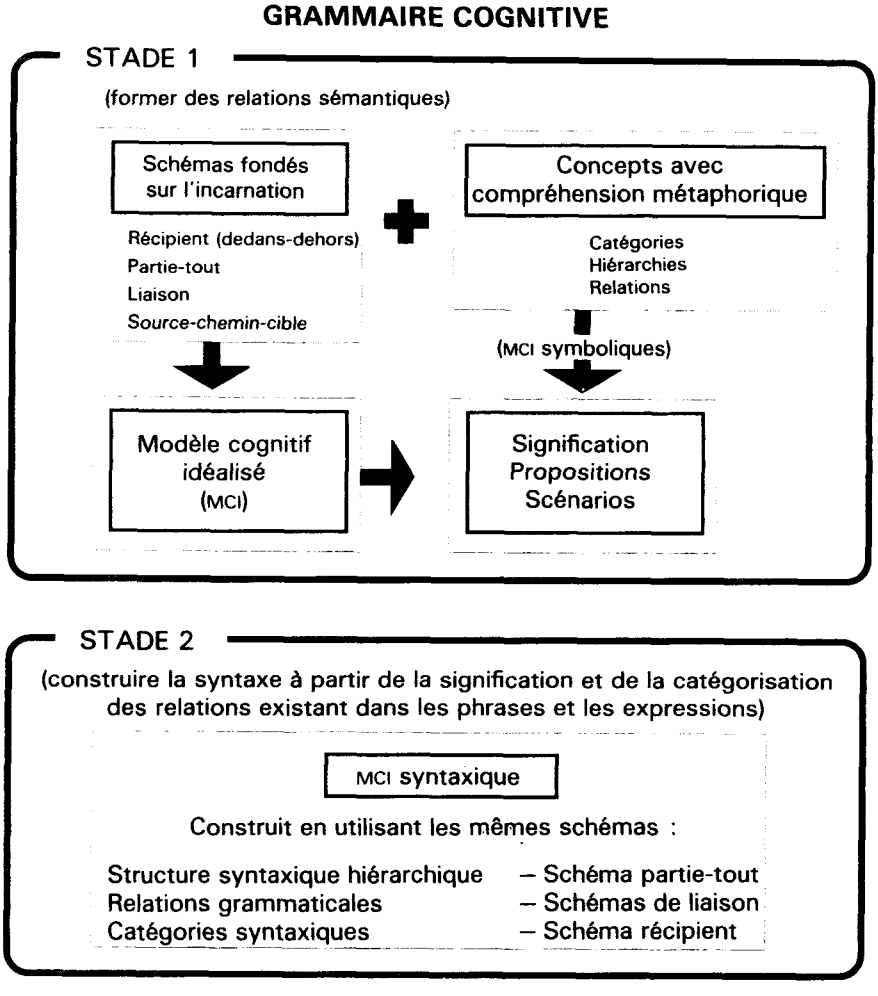


FIGURE P-8

*Exemple des processus d'une grammaire cognitive d'après Lakoff. Contrairement aux grammaires génératives (voir la figure P-7), les règles sont acquises à travers l'expérience linguistique et la signification naît de l'incarnation des concepts. Il n'a pas encore été totalement démontré que ce type de grammaire possède la puissance analytique des grammaires génératives conventionnelles telles que la grammaire fonctionnelle lexicale de Bresnan. Mais la grammaire cognitive fournit néanmoins une façon de relier la signification (à travers l'incarnation) à la catégorisation et à la structure des phrases. Les différents « stades » ne surviennent pas nécessairement de façon séquentielle dans le temps ; ils se recouvrent partiellement. Bien entendu, leurs intersections sont minimales au début de l'acquisition du langage. (*MCI *: modèle cognitif idéalisé.)*

de l'expérience qui donne lieu aux schémas. Le schéma *source-chemin-cible*, par exemple, naît de notre fonctionnement corporel, envahit notre expérience, il est bien structuré et il est bien compris. Le domaine source et le domaine cible de toute métaphore fondée sur ce schéma seront corrélés par l'expérience selon ce schéma. Les concepts préalables de niveau de base et les concepts des schémas d'images possèdent une signification immédiate et constituent la base de ce schéma. Ils fournissent également des points de départ aux règles de composition sémantique qui permettent d'obtenir des concepts plus complexes à partir de concepts plus simples.

Ces idées sont encapsulées dans « l'hypothèse de la spatialisation de la forme » de Lakoff. D'après cette hypothèse, les catégories se comprennent sous l'angle des schémas *récipients*, la structure hiérarchique sous l'angle des schémas *partie-tout* et *haut-bas*, la structure relationnelle sous l'angle des schémas de *liaison*, la structure radiale des catégories sous l'angle des schémas *centre-périphérie*, la structure avant-plan/arrière-plan sous l'angle des schémas *avant-arrière*, et les échelles de quantités linéaires sous l'angle des schémas *haut-bas* et des schémas *d'ordre linéaire*. Toutes font appel à une mise en correspondance métaphorique de structures physiques (ou spatiales) avec des structures conceptuelles.

Mais où intervient au juste le langage ? Dans les *modèles cognitifs idéalisés*, dont certains sont des structures faites de symboles. Ces modèles sont de l'un des cinq types suivants : modèles de schémas d'images, métaphoriques, métonymiques, propositionnels et symboliques. Parmi eux, ceux qui aboutissent à la fonction linguistique sont les modèles cognitifs idéalisés propositionnels et symboliques.

Les modèles cognitifs idéalisés propositionnels n'utilisent ni la métaphore, ni la métonymie, ni les images mentales. En revanche, ils utilisent des concepts de niveau de base – entités, actions, états et propriétés. Les propositions simples suivent le schéma partie-tout : la proposition constitue le tout, le prédicat étant l'une de ses parties et les arguments (l'agent, le patient, celui qui ressent, l'instrument, le lieu, etc.) l'autre. Les relations sémantiques sont construites à partir de schémas de liaison, et les propositions complexes sont ensuite formées à partir des propositions simples par modification, quantification, conjonction, négation, et ainsi de suite. De plus, il est possible de construire des scénarios d'un état initial, d'une suite d'événements et d'un état final, structurés par un schéma *source-chemin-cible*.

Lorsque des éléments linguistiques sont associés à des modèles

cognitifs idéalisés *conceptuels*, ces modèles deviennent des modèles cognitifs idéalisés *symboliques*. Ils peuvent ensuite être caractérisés en fonction des morphèmes et des mots de chaque langue particulière. Les noms, par exemple, sont des catégories radiales (les catégories centrales sont les gens, les lieux, les choses ; les catégories non centrales sont les noms abstraits tels que « force »). Les *verbes* sont également des catégories radiales (les catégories centrales sont les actions physiques de niveau de base telles que courir, frapper, donner). Les membres restants de ces catégories sont motivés par leurs relations avec ces membres centraux. Le lien avec la signification est évident.

Que dire de la syntaxe elle-même ? Lakoff affirme que les principes dont il parle fournissent des bases sémantiques aux catégories syntaxiques. D'après sa théorie, la structure syntaxique hiérarchisée (voir un exemple sur la figure P-7) est elle-même caractérisée par des schémas *partie-tout*, les structures de tête et les structures modificatives sont caractérisées par des schémas *centre-périphérie*, les relations grammaticales sont caractérisées par des schémas de *liaison* et les catégories syntaxiques sont caractérisées par des schémas *récipients*. Remarquez ce qui se passe ici : *les constructions grammaticales sont elles-mêmes des modèles cognitifs idéalisés*. Ainsi, l'appariement de la signification et de la syntaxe revient à l'appariement direct d'un modèle cognitif idéalisé de la syntaxe et d'un modèle cognitif idéalisé *préalable* de la signification. Les régularités de la structure grammaticale et celles du lexique peuvent être décrites en termes de catégories radiales, et les mots ayant plusieurs sens s'expliquent également en ces termes.

Selon ce point de vue et celui de Langacker, le langage est fondé sur la cognition – c'est-à-dire sur des modèles cognitifs qui peuvent s'expliquer par rapport au fonctionnement de l'organisme. Cette base cognitive est soumise aux contraintes imposées par la nature de la réalité physique, et elle dépend aussi de l'imagination et des interactions sociales. La signification découle de l'incarnation et de la fonction, la compréhension naît lorsque les concepts ont une signification dans ce sens, et on considère que la vérité apparaît lorsque la compréhension d'un énoncé par un individu correspond suffisamment à sa compréhension d'une situation pour lui permettre de satisfaire ses objectifs personnels. (Remarquez le pragmatisme !) Il n'y a donc pas de vérité absolue, pas de vue de l'extérieur. Notre vision de ce qui existe (la métaphysique) n'est pas indépendante de la façon dont nous sommes venus à en avoir connaissance (l'épis-

témologie). Comme le dit Lakoff : « La vérité est une opération de *bootstrapping*, fondée sur des liens directs avec une expérience structurée de manière préconceptuelle et claire, et sur les concepts qui s'accordent avec cette expérience. » Cela correspond aux propositions relatives au réalisme pondéré que j'ai formulées au chapitre 15.

La connaissance, tout comme la vérité, est un concept radial. Elle dépend de notre compréhension, des concepts de niveau de base, et aussi d'une compréhension socialement admise. Elle n'est solide que dans la mesure où la compréhension humaine peut l'être, et elle est toujours sujette à révision. L'objectivité n'est pas absolue ; au contraire, elle suppose que l'on considère une situation donnée sous autant d'angles que possible, et que l'on distingue les concepts de niveau de base et les concepts des schémas d'images de ceux qui n'ont qu'une signification indirecte.

Il est évident que l'exemple fourni par la grammaire cognitive de Lakoff (figure P-8) diffère radicalement des grammaires génératives plus généralement admises (voir la figure P-7). Elle en diffère par la philosophie, le style et la méthodologie. Elle s'accorde davantage avec les bases biologiques des fonctions cérébrales et corporelles, et avec les données psychologiques sur la catégorisation. Elle permet d'éviter les erreurs de catégorie suscitées par la notion de « langage de la pensée » et l'erreur objectiviste inhérente à la grammaire générative. Il s'agit d'une proposition importante et pleine d'imagination. Mais, bien qu'elle suggère que l'incarnation est à l'origine de la signification, elle ne fournit pas d'explications sur la *façon* dont cela pourrait se produire. Et elle n'indique pas non plus comment les modèles cognitifs idéalisés du langage résultent des mécanismes de catégorisation perceptive et conceptuelle. Pour cela, il faut disposer d'une théorie biologique générale de la fonction cérébrale et d'une théorie de la conscience, fondées toutes deux sur les données de l'évolution et du développement. Ce sont ces théories que j'ai essayé de construire au cours de ma trilogie et de récapituler dans le présent volume.

Il peut être utile de faire quelques commentaires à propos de la relation entre la grammaire cognitive de Lakoff et la théorie de l'acquisition du langage décrite au chapitre 12. La grammaire cognitive est fondée sur la notion d'incarnation, mais elle ne spécifie pas comment se produit l'incarnation. Elle recherche plutôt des indices du fait que les catégories radiales, les métaphores et les métonymies constituent des structures orientatrices du langage. Et de même, elle fait appel à la catégorisation pour rendre compte de l'apparition

des relations syntaxiques. À tous ces égards, elle est compatible avec la théorie épigénétique présentée au chapitre 12. Mais cette théorie permet de clarifier les questions liées à l'évolution et à l'acquisition du langage d'une manière dont la théorie de Lakoff, à qui il manque une description des mécanismes de l'incarnation, s'avère incapable. En fait, la théorie épigénétique fournit des raisons supplémentaires de prendre en compte certains aspects d'une théorie générative *structurelle* approfondie de la grammaire, comme celle de Bresnan (qui met l'accent sur le lexique), et de les relier à une théorie *fondée sur les catégories* comme celle de Lakoff. L'approche de Langacker, tout en rejetant les aspects génératifs de celle de Bresnan, lui ressemble parce qu'il met également l'accent sur l'importance du lexique. Mais, pour comprendre pleinement les formulations grammaticales, on doit analyser les mécanismes cérébraux de formation des concepts, de formation des valeurs-catégories, de liaison avec le phénotype et de liaison avec les mécanismes de la conscience. Et on doit également explorer les grammaires des langues particulières dans les termes décrits par Langacker, par Bresnan et d'autres. Un riche domaine d'étude pourrait s'ouvrir si l'on essayait de découvrir comment une théorie de l'incarnation telle que celle que j'ai présentée pourrait permettre de rapprocher et d'associer ces tentatives importantes, mais différentes, de construction d'une théorie linguistique.

Le livre de Lakoff – *Women, Fire and Dangerous Things* – est sorti à peu près en même temps que mon livre *Neural Darwinism*, dans lequel j'essayais de jeter les bases d'une théorie globale du cerveau. Je sais qu'à l'époque, je n'étais pas au courant de la sortie de son livre, et je suppose qu'il n'était pas non plus au courant de celle du mien. Le problème central abordé dans *Neural Darwinism* était la catégorisation perceptive. Dans l'un de mes ouvrages intitulé *The Remembered Present : a Biological Theory of Consciousness*, j'ai généralisé la théorie du cerveau à l'expérience perceptive, à la formation des concepts et au langage. Rétrospectivement, il semble que ces deux livres complètent tout à fait les travaux de Langacker, de Lakoff et de Johnson, car ils fournissent une base biologique essentielle à un grand nombre de leurs propositions concernant l'importance de l'incarnation pour la grammaire et la cognition. Mais ni leurs travaux ni le mien ne récusent l'importance des efforts accomplis par d'autres linguistes pour tenter de comprendre la structure syntaxique. L'importance de leurs efforts et de ceux des psychologues cogniticiens est énorme. Il n'en reste

pas moins que dans la biologie, ces efforts demeureront insuffisants et induiront même parfois en erreur. C'est cela que j'ai essayé de montrer dans cette postface.

J'espère que ceux qui auront lu à la fois le livre et la postface auront compris suffisamment clairement le défi posé : nous devons incorporer la biologie dans nos théories de la connaissance et du langage. Pour ce faire, nous devons mettre au point ce que j'ai appelé une épistémologie fondée sur la biologie – c'est-à-dire expliquer comment nous faisons pour savoir et pour être conscients, à la lumière des données de l'évolution et de la biologie du développement. Si nous parvenons à atteindre plus pleinement cet objectif, nous élargirons énormément nos horizons scientifiques. En nous permettant de comprendre ce qui fait de nous des êtres humains, cette épistémologie enrichira nos vies.

Bibliographie

J'ai réuni les références bibliographiques suivantes à l'intention des lecteurs qui voudraient obtenir de plus amples informations de base ou approfondir certaines idées. Pour chaque travail original récent évoqué au cours du livre, j'ai essayé de citer un article pertinent afin de respecter – ne serait-ce que de façon minimale – les normes des publications académiques.

Pour une bibliographie plus détaillée, le lecteur pourra consulter les références figurant dans ma trilogie sur la morphologie et l'esprit : *Topobiology*, *Neural Darwinism* et *The Remembered Present* (ces trois livres sont publiés par Basic Books et figurent ci-dessous).

Les remarques qui accompagnent les références peuvent être utiles pour certains. Mais elles reflètent mon opinion personnelle et ne se veulent pas du tout exhaustives.

CHAPITRE 1

JAMES, W., *The Principles of Psychology*, 1890. Réédition, Dover (New York), 1950. (Traduction française, en version abrégée, sous le titre *Précis de psychologie*, éditions Marcel Rivière, 1921.)

Une œuvre monumentale, par l'un des fondateurs de la psychologie expérimentale : elle comporte des descriptions et des analyses très pénétrantes, ainsi que des opinions personnelles très marquées.

James, W., « Does consciousness exist ? », *in The Writings of William James*, J. J. McDermott (ed.), University of Chicago Press (Chicago), 1977, pp. 169-183.

Le travail d'un pionnier, qui montre que la conscience est un processus et non une chose ou une substance.

FLANAGAN, O. J., Jr., *The Science of the Mind* (2e éd.), MIT Press (Cambridge, Mass.), 1991.

Très bonne vue d'ensemble des réflexions et du travail des psychologues modernes, avec une évaluation équilibrée de leur importance actuelle.

BRENTANO, F., *Psychology from an Empirical Standpoint*, O. Kraus et L. L. McAlister (eds), A. C. Rancurello *et al.* (trad.), Humanities (Highlands, N.J.), 1973.

L'œuvre clé de ce psychologue, philosophe et ancien prêtre qui mit l'accent sur l'importance de l'intentionnalité. Professeur à Vienne, Brentano influença la pensée de Freud, qui assista à ses conférences.

GREGORY, R. L., *Mind in Science*, Cambridge University Press (Cambrige), 1981.

Compte rendu historique des difficiles rapports existant entre la méthodologie scientifique et la question de l'esprit. Décousu, peu concluant, mais riche en suggestions.

GRIFFIN, D. R., *La Pensée animale*, Denoël, 1988.

Fougueux arguments en faveur de l'existence d'une conscience chez l'animal, sous la plume d'un éthologue de renom. Dans d'autres ouvrages, l'auteur a suggéré que même les abeilles pourraient être conscientes. Je pense que les arguments présentés sont insuffisants (et en fait, que l'existence d'une telle conscience est extrêmement peu probable), mais le point de vue excentrique de Griffin est utile par son côté provocateur.

PREMACK, D. et PREMACK A. J., *L'Esprit de Sarah*, Fayard (Le Temps des sciences), 1984.

Description splendide et extrêmement claire des capacités mentales des chimpanzés. Actuel, érudit et stimulant.

MCCULLOCH, W. S., *Embodiments of Mind*, MIT Press (Cambridge, Mass.), 1989.

Recueil d'essais, écrits par un brillant précurseur de la neurobiologie moderne. Réédition. Utile et plein d'imagination.

BLAKEMORE, C. et GREENFIELD, S. (eds.), *Mindwaves : Thoughts on Intelligence, Identity, and Consciousness*, Blackwell (New York), 1987.

Recueil d'articles sur ces thèmes par des neurobiologistes et des philosophes. Donne un aperçu des principaux problèmes, des malentendus, et les différents points de vue de divers praticiens.

GREGORY, R. L. (ed.), *The Oxford Companion to the Mind*, Oxford University Press (Oxford), 1987.

Petite encyclopédie comportant des articles sur des questions très variées, écrites par de nombreux experts. Incomplet, mais utile et très amusant pour ceux qui ont de la curiosité intellectuelle. Très agréable à feuilleter.

CHAPITRE 2

WHITEHEAD, A. N., *Science and the Modern World*, Macmillan (New York), 1925.

Description classique, faite par un logicien, historien des sciences et

métaphysicien. Place la relation existant entre l'observateur scientifique et l'individu réfléchi et subjectif dans une riche perspective historique.

GALILEI, G., « The Assayer », 1623, S. Drake (trad.), *in Discoveries and Opinions of Galileo*, Doubleday (New York), 1957.

DRAKE, S., *Galilée*, Actes Sud, 1987.

Dans ces deux livres, le « fondateur » est présenté, et sa pensée résumée, par l'un de ses plus dévoués historiens. Je conseille à ceux qui sentent que les chercheurs actuels ne pensent qu'aux priorités de lire les doléances de Galilée au début de « The Assayer » questions de « paternité » des résultats, (« L'expérimentateur »). Disons également au passage que Galilée avait compris la nature des qualités secondaires (couleur, chaleur, etc.) presque un siècle avant Locke.

DESCARTES, R., *Le Discours de la méthode*, Flammarion, 1966.

DESCARTES, R., *Les Méditations métaphysiques*, Flammarion, 1979 et LGF (Le Livre de poche – Classiques de la philosophie), 1990.

DESCARTES, R., *Les Passions de l'âme*, Gallimard, 1988 et LGF (Le Livre de poche – Classiques de la philosophie), 1990.

Si Galilée est le fondateur de la science moderne, Descartes est celui de la philosophie moderne. Ses pensées démontrent que le génie, même lorsqu'il mène à des conclusions fausses, peut continuer à avoir une influence majeure pendant très longtemps. Nous nous débattons encore avec les questions posées par Descartes.

CHAPITRE 3

SHEPHERD, G., *Neurobiology*, Oxford University Press (New York), 1983.

Description élémentaire standard des résultats de la neurobiologie moderne. Il en existe beaucoup d'autres, mais celle-ci contient la plupart des notions de base et aborde un certain nombre de questions au niveau des principes.

LURIA, A. R., *The Working Brain : An Introduction to Neuropsychology*, Basic Books (New York), 1973.

Ce grand clinicien et neurologue soviétique, aujourd'hui décédé, donne un aperçu très clair de ce qui se passe lorsque des parties du cerveau sont perturbées. En suivant le cours de ses descriptions, le néophyte parvient à se faire une meilleure idée des fonctions du cerveau proprement dit qu'avec un livre comme celui de Shepherd.

DIDEROT, D., *Le Rêve de d'Alembert*, et *Le Neveu de Rameau*, Messidor – Éditions sociales (Essentiel), 1984 et LGF (Le Livre de poche – Classique), 1984.

Ce petit classique est cité dans le texte. L'original se trouve à Saint-Pétersbourg, car Diderot fut conseiller de Catherine la Grande de

Russie. Si j'ai bien compris, d'Alembert, collaborateur de Diderot dans la réalisation de l'*Encyclopédie*, ne fut pas du tout content de voir sa relation avec Mademoiselle de l'Espinasse ainsi étalée au grand jour.

CHANGEUX, J.-P., *L'Homme neuronal*, Fayard (Le Temps de sciences), 1983 et Hachette-Pluriel, 1984.

Ouvrage grand public écrit par un neurobiologiste qui, comme moi, pense que le cerveau est un système sélectif. Le livre contient de brefs aperçus des questions, aborde les aspects historiques et donne une description sommaire des liens entre psychologie cognitive et neurophysiologie.

CHAPITRE 4

À ceux qui seraient intéressés par une histoire élémentaire de la philosophie, je conseille B. RUSSELL, *A History of Western Philosophy*, Simon and Schuster (New York), 1945. Il s'agit là d'un ouvrage clair, plein de préjugés et très stimulant. Pour ceux qui préfèrent un ouvrage plus récent et plus rigoureux, voir A. J. AYER, *Philosophy in the Twentieth Century*, Vintage (East Hanover, N. J.), 1984. On trouvera également un bon ensemble de comptes rendus historiques de psychologie moderne (un peu techniques, toutefois) dans E. HEARST (ed.), *The First Century of Experimental Psychology*, Lawrence Erlbaum Associates (Hillsdale, N. J.), 1979.

KANIZSA, G., *Organization in Vision : Essays on Gestalt Perception*, Praeger (New York), 1979.

Description merveilleusement révélatrice du monde de ce qu'on appelle les illusions visuelles. Une œuvre de maître, un petit classique, superbement illustré.

CHAPITRE 5

DARWIN, C., *L'Origine des espèces*, Découverte (Fondations), 1985.

Le chef-d'œuvre qui a jeté les bases de la biologie moderne.

BARRETT, P. H., GAUTREY, P. J., HERBERT, S., KOHN, D. et SMITH, S. (eds), *Charles Darwin's Notebooks, 1836-1844 : Geology, Transmutation of Species, Metaphysical Enquiries*, Cornell University Press (Ithaca, N. Y.), 1987.

Comment un grand scientifique et penseur voyait l'esprit.

ROMANES, G. J., *Mental Evolution in Animals*, Appleton (New York), 1884.

ROMANES, G. J., *Mental Evolution in Man*, Appleton (New York), 1889.

Les réflexions d'un contemporain de Darwin. De bons exemples de la façon dont une théorie remarquable et provocatrice devient très répandue.

MAYR, E., *Histoire de la biologie : diversité, évolution et hérédité*, Fayard (Le Temps des sciences), 1989.
Un chef-d'œuvre moderne par un grand évolutionniste. L'une des meilleures descriptions de Darwin, du darwinisme et des « sous-théories » qui entrent dans la composition de cette théorie extrêmement complexe qu'est la théorie moderne de l'évolution.
RICHARDS, R. J., *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*, University of Chicago Press (Chicago), 1987.
Exposé très complet et très érudit. Le meilleur en date sur ce thème.

CHAPITRE 6

THOMPSON, D. W., *On Growth and Form*, Cambridge University Press (Cambridge), 1942.
L'un des grands classiques sur le thème de la forme animale, par un auteur plein de talent ne croyant pas au darwinisme. Bien que ses exemples directs n'aient pas grand-chose à voir avec le système nerveux, l'ouvrage n'en est pas moins fascinant.
EDELMAN, G. M., *Topobiology : An Introduction to Molecular Embryology*, Basic Books (New York), 1988.
Compte rendu plus détaillé et fondamental du thème de ce chapitre. Le dernier chapitre de *Topobiology*, qui, en vertu du sujet qu'il aborde, constitue le premier tome de la trilogie sur la morphologie et l'esprit (bien qu'il n'ait pas été publié en premier), décrit les liens entre la topobiologie et les théories sélectionnistes du cerveau.

CHAPITRE 8

BURNET, F. M., *The Clonal Selection Theory of Acquired Immunity*, Vanderbilt University Press (Nashville), 1959.
Il s'agit là de la première description détaillée des conceptions sélectionnistes de Burnet. Voir le livre de Mayr cité ci-dessus pour la toile de fond évolutionniste.

CHAPITRE 9

EDELMAN, G. M., *Neural Darwinism : The Theory of Neuronal Group Selection*, Basic Books (New York), 1987.
Ce livre expose la théorie de la sélection des groupes neuronaux *in extenso*, avec tous les arguments à l'appui. Beaucoup plus érudit que le présent ouvrage. On trouvera le premier exposé schématique de la

théorie dans Edelman, G. M. et Mountcastle, V. B., *The Mindful Brain*, MIT Press (Cambridge, Mass.), 1978.

BARLOW, H. B., « Neuroscience : a new era ? », *Nature*, *331*, p. 571, 1988.

CRICK, F., « Neural edelmanism », *Trends in Neuroscience*, *12*, pp. 240-248, juillet 1989.

PURVES, D., *Body and Brain : A Trophic Theory of Neural Connections*, Harvard University Press (Cambridge, Mass.), 1988.

Ces trois auteurs critiquent divers aspects de la théorie de la sélection des groupes neuronaux. Les contre-attaques sont brièvement présentées dans ce chapitre. Vive le sport ! [1].

MICHOD, R. E., « Darwinian selection in the brain », *Evolution*, *43*, n° 3, pp. 694-696, 1989.

Un bilan en faveur de l'idée que la théorie de la sélection des groupes neuronaux constitue une description sélectionniste dans la droite ligne du raisonnement en termes de populations. Voir également sa réponse aux critiques de la TSGN émises par Crick, accompagnée de la réponse de Reeke, G. N., Jr., dans *Trends in Neuroscience*, *13*, n° 1, pp. 11-14, 1990.

Pour de plus amples informations sur le travail d'Eckhorn et de ses collègues, et sur celui de Gray et des siens, voir la bibliographie dans Sporns, O., Gally, J. A., Reeke, G. N., Jr. et Edelman, G. M., « Reentrant signaling among stimulated neuronal groups leads to coherency in their oscillatory activity », *Proceedings of the National Academy of Sciences*, *86*, pp. 7265-7269, 1989.

CHAPITRES 10 À 13

EDELMAN, G. M., *The Remembered Present : A Biological Theory of Consciousness*, Basic Books (New York), 1989.

Dernier tome de la trilogie sur la morphologie et l'esprit. Tente de fournir une description scientifique des bases de la conscience fondée sur des principes. Bien qu'il constitue le dernier volume de la trilogie, on m'a déjà dit qu'il gagne à être lu en premier.

MARCEL, A. J. et BISIACH, E. (eds), *Consciousness in Contemporary Science*, Clarendon (Oxford), 1988.

Recueil d'articles de qualité sur ce thème, englobant un très large domaine.

BARTLETT, F. C., *Remembering : A Study in Experimental and Social Psychology*, Cambridge University Press (Cambridge), 1964.

BARTLETT, F. C., *Thinking : A Study of Humans*, Basic Books (New York), 1958.

---

1. En français dans le texte (N.d.T.).

Deux classiques, surtout le premier qui donne une analyse profonde de l'acte de se souvenir.

HEBB, D. O., *The Organization of Behavior : A Neuropsychological Theory*, Wiley (New York), 1949.

HEBB, D. O., *Essay on Mind*, Lawrence Erlbaum Associates (Hillsdale, N. J.), 1980.

Le premier est l'une des premières tentatives visant à rendre compte des phénomènes psychologiques en termes d'interactions neuronales et cellulaires. Le second contient les réflexions plus tardives de ce maître de la science moderne.

FREUD, S., « Project for a Scientific Psychology », *in The Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol. 1, pp. 283-411, Strachey, J. (ed.), Hogarth (Londres), 1976.

FREUD, S., *Introduction à la psychanalyse*, Payot (Prisme et Petite Bibliothèque), 1989.

FREUD, S., *Sur la psychanalyse : cinq conférences*, Gallimard (Connaissance de l'inconscient – Nouv. trad.), 1991.

FREUD, S., *Nouvelles conférences d'introduction à la psychanalyse*, Gallimard (Connaissance de l'inconscient), 1984.

FREUD, S., *Nouvelles conférences sur la psychanalyse*, Gallimard (Folio-Essais), 1989.

FREUD, S., *Sur le rêve*, Gallimard (Connaissance de l'inconscient – Nouv. trad.), 1988 et Gallimard (Folio-Essais) 1990.

L'« Esquisse de psychologie scientifique » commença par ravir Freud et finit par lui faire horreur. S'il est parvenu jusqu'à nous, c'est parce qu'il a été préservé pour la postérité par son amie Marie Bonaparte. Le reste de ces œuvres représente les meilleurs éléments du travail du maître au niveau grand public. Freud considérait le dernier – un chef-d'œuvre – comme son meilleur travail.

ERDELYI, M. H., *Psychoanalysis : Freud's Cognitive Psychology*, Freeman (New York), 1985.

Excellent exposé des concepts freudiens clés, et en particulier des idées de Freud sur la mémoire et l'inconscient.

BROUWER, L. E. J., « Consciousness, philosophy, and mathematics », *in Proceedings of the Tenth International Congress of Philosophy*, vol. 2, pp. 1235-1249, Beth, E. W., Pos, H. J., et Hollak, J. H. A. (eds), North Holland (Amsterdam), 1949.

Remarquable travail d'imagination par un spécialiste de topologie et de philosophie des mathématiques. Post-kantien, ardu, mais très stimulant.

HILGARD, E. R., *Divided Consciousness : Multiple Controls in Human Thought and Action* (éd. augmentée), Wiley (New York), 1977.

Une conception différente, résolument post-freudienne. Plein d'exemples fascinants et de réflexions concernant les dédoublements de la cons-

cience et les phénomènes hypnotiques. Ceux qui seraient intéressés par une exploration étonnante des phénomènes gestaltistes visuels pourront consulter le merveilleux livre de Gaetano Kanizsa, cité dans la liste correspondant au chapitre 4.

STADDON, J. E. R., *Adaptive Behavior and Learning*, Cambridge University Press (Cambridge), 1983.

Un bon exposé des questions d'apprentissage vues sous l'angle de la biologie en général.

ALEXANDER, R. D., « Evolution of the human psyche », *in The Human Revolution*, Mellars, P. et Stringer, C. (eds), Princeton University Press (Princeton, N. J.), 1989.

L'auto-duperie en tant que phénomène adaptatif permettant de duper les autres au cours de la lutte pour la survie – une hypothèse peu plausible mais fascinante.

CHAPITRE 14

ROTA, G. C., « Mathematics and philosophy : the story of a misunderstanding », *Review of Metaphysics*, *44*, pp. 259-271, décembre 1990.

Un article réfléchi, écrit de façon caustique par un éminent mathématicien, sur la méfiance dont il faut faire preuve à l'égard de l'axiomatisation à outrance. « Le *symbol dropping* [1] snobinard que l'on trouve aujourd'hui dans les articles de philosophie, écrit notamment l'auteur, suscite la méfiance des mathématiciens. C'est comme si vous alliez chez l'épicier et que vous y trouviez quelqu'un en train d'essayer de régler sa note avec l'argent d'un jeu de Monopoly ».

CHAPITRES 15 ET 16

EDWARDS, P. (ed.), *The Encyclopedia of Philosophy*, vols. 1-4, The Free Press (New York), 1973.

DANTO, A., *Connections to the World : The Basic Concepts of Philosophy*, Harper & Row (New York), 1989.

RUSSELL, B., *Les Problèmes de philosophie,* Payot (Paris), 1990.

Une œuvre de référence et deux introductions pleines de lucidité, l'une récente, l'autre ancienne.

BECHTEL, W., *Philosophy of Mind : An Overview for Cognitive Science*, Lawrence Erlbaum Associates (Hillsdale, N. J.), 1988.

CHURCHLAND, P. M., *Matter and Consciousness*, MIT Press (Cambridge, Mass.), 1984.

1. *Symbol dropping* : jeu de mots sur l'expression « name dropping », qui signifie tendance à truffer son discours des noms de gens en vue qu'on connaît (N.d.T.).

Deux courtes introductions à la philosophie de l'esprit.

QUINE, W. V., *Quiddities : An Intermittently Philosophical Dictionary*, Harvard University/Belknap Press (Cambridge, Mass.), 1987.

Recueil de notes amusantes et personnelles, par un éminent philosophe et logicien américain.

WITTGENSTEIN, L., *Investigations philosophiques*, Gallimard (Tel), 1986.

Édition posthume des conceptions révisionnistes de l'un des esprits philosophiques les plus intéressants de ce siècle. A trait à la question des catégorisations, abordée tout au long du présent ouvrage.

WHITEHEAD, A. N., *Modes of Thought*, The Free Press (New York), 1938.

Réflexions métaphysiques « tout court », d'un penseur moderne provisoirement tombé en disgrâce dans la plupart des milieux universitaires actuels. Vaut vraiment la peine d'être lu pour les réflexions suggestives et pleines d'imagination qu'il contient.

NAGEL, T., « What is it like to be a bat ? », *Philosophical Review*, *83*, pp. 435-450, 1974.

NAGEL, T., *The View from Nowhere*, Cambridge University Press (New York), 1986.

Analyses claires et pénétrantes des dilemmes de l'épistémologie et de la métaphysique.

RYLE, G., *La notion d'esprit : pour une critique des concepts mentaux*, Payot (Bibliothèque scientifique), 1978.

Critique cinglante des erreurs de catégorie commises en philosophie de l'esprit, par l'inventeur de l'expression « le fantôme dans la machine ».

RUSSELL, B., *A History of Western Philosophy*, Simon & Schuster (New York), 1945.

Déjà cité. Lucide, *sui generis* ; écrit par l'un des pionniers de la logique mathématique – et l'un des philosophes modernes aux opinions les plus courageuses.

AYER, A. J., *The Problem of Knowledge*, Penguin (Middlesex, N. J.), 1956.

Vue d'ensemble épistémologique de la question par un ancien adepte de l'empirisme logique.

PIAGET, J., *Biology and Knowledge : An Essay on the Relations Between Organic Relations and Cognitive Processes*, University of Chicago Press (Chicago), 1971.

La vision, présentée ici à titre de comparaison, d'un grand psychologue du développement. Non seulement très personnel et original, mais également révélateur de l'abîme qui sépare l'attitude des scientifiques de celle des philosophes. Plutôt métaphorique dans ses comparaisons de l'embryologie et de la psychologie.

DAVIS, P. J. et HERSH, R., *Descartes's Dream : The World According to Mathematics*, Houghton Mifflin (Boston), 1986.

Une très belle œuvre, écrite par deux mathématiciens, sur la nature et

les limites des mathématiques. Les auteurs n'ont aucune sympathie à l'égard de la vision platonicienne des mathématiques.

MORGAN, M. J., *Molyneux's Question : Vision, Touch, and the Philosophy of Perception*, Cambridge University Press (Cambridge), 1977.

Un élégant essai sur certains aspects historiques de la psychologie de la perception spatiale. Peu importe le point de vue de la chauve-souris – que se passerait-il si vous aviez toujours été aveugle et que vous retrouviez soudain la vue ? Votre « espace tactile » correspondrait-il à votre « espace visuel » ?

HULL, J. M., *Touching the Rock : An Experience of Blindness*, Pantheon (New York), 1990.

Émouvante description de la manière dont la conscience d'un individu est modifiée par la cécité.

BOYD, R. et RICHERSON, P. J., *Culture and the Evolutionary Process*, University of Chicago Press (Chicago), 1985.

Description remarquablement bien équilibrée de la façon dont le comportement social humain et l'évolution pourraient interagir. L'une des meilleures explorations de ce dangereux terrain.

BARROW, J. D. et TIPLER, F. J., *The Anthropic Cosmological Principle*, Oxford University Press (Oxford), 1988.

La source de ma citation sur les corbeilles à papier. Peut-être que le nombre relativement élevé de références à la philosophie qui apparaissent dans cette section sert à prouver ce qui est dit dans cette citation (voir page 207).

FLEW, A., *An Introduction to Western Philosophy : Ideas and Arguments from Plato to Popper*, Thames and Hudson (New York), 1989.

Contient une bonne discussion de l'idée de l'âme.

Voir aussi les références du chapitre 4.

CHAPITRE 17

ARENDT, H., *La Vie de l'esprit*, vol. 1 : *La pensée* (2e éd.), PUF (Philosophie d'aujourd'hui), 1987 ; vol. 2 : *La volonté*, PUF (Philosophie d'aujourd'hui), 1983.

Abrégé de philosophie. Il est révélateur de comparer le point de vue présenté dans cet ouvrage à ceux d'expérimentateurs comme Bartlett (pour des références à Bartlett, voir la bibliographie des chapitres 10-13).

LANGER, S. K., *Mind : An Essay on Human Feeling*, vol. 1-3, Johns Hopkins University Press (Baltimore), 1967, 1972, 1973.

Vue d'ensemble magistrale, écrite par une philosophe historienne des idées et spécialiste du symbolisme artistique. Malheureusement, l'écriture du troisième volume de cette œuvre dut être interrompue en raison de la cécité progressive de l'auteur.

MANDLER, G., *Mind and Body : Psychology of Emotion and Stress*, Norton (New York), 1984.

SOLOMON, R. C., *The Passions*, University of Notre Dame Press (Notre Dame), 1983.

DE SOUSA, R., *The Rationality of Emotion*, MIT Press (Cambridge, Mass.), 1987.

Trois exposés sur les émotions, le premier scientifique et les deux autres philosophiques. Ensemble, ils font ressortir les différents niveaux et la nature extraordinairement complexe des émotions.

CHAPITRE 18

WILLIAMS, M., *Brain Damage, Behavior, and the Mind*, Wiley (New York), 1979.

KOLB, B. et WHISHAW, I. Q., *Fundamentals of Human Neuropsychology* (3e éd.), Freeman (San Francisco), 1990.

MCCARTHY, R. A. et WARRINGTON, E. K., *Cognitive Neuropsychology : A Clinical Introduction*, Academic (New York), 1990.

Trois livres sur les effets des lésions cérébrales (voir également A. Luria, *The Working Brain*, parmi les références du chapitre 1).

KAPLAN, H. I. et SADOCK, B., *Comprehensive Textbook of Psychiatry/IV*, vol. 1-2, Williams & Wilkins (Baltimore), 1989.

Un gros texte de psychiatrie. Pour les plus courageux.

MODELL, A. H., *Other Times, Other Realities : Towards a Theory of Psychoanalytic Treatment*, Harvard University Press (Cambridge, Mass.), 1990.

HUNDERT, E. M., *Philosophy, Psychology, and Neuroscience : Three Approaches to the Mind*, Clarendon (Oxford), 1989.

Deux psychiatres appliquent la théorie de la sélection des groupes neuronaux à des aspects de leur sujet d'étude.

SCHACTER, D. L., MCANDREWS, M. P. et MOSCOVITCH, M., « Access to consciousness : dissociations between implicit and explicit knowledge in neuropsychological syndromes », *in Thought Without Language*, Weiskrantz, L. (ed.), pp. 242-278, Clarendon (Oxford), 1988.

BISIACH, E., « Language without thought », *in Thought Without Language*, Weiskrantz, L. (ed.), pp. 465-491, Clarendon (Oxford), 1988.

Deux articles révélateurs sur certains syndromes dissociatifs de la conscience.

SACKS, O., *L'Homme qui prenait sa femme pour un chapeau*, Seuil, 1988.

Un fascinant ensemble de « contes », par un clinicien très humain qui est également un extraordinaire conteur.

CHAPITRE 19

REEKE, G. N., Jr. et EDELMAN, G. M., « Real brains and artificial intelligence », *Daedalus*, *117*, n° 1, pp. 143-173, Hiver 1988.

REEKE, G. N., Jr., FINKEL, L. H., SPORNS, O., et EDELMAN, G. M., « Synthetic neural modeling : a multilevel approach to the analysis of brain complexity », *in Signal and Sense : Local and Global Order in Perceptual Maps*, EDELMAN, G. M., GALL, W. E. et COWAN, W. M. (eds), pp. 607-107, Wiley-Liss (New York), 1990.

EDELMAN, G. M. et REEKE, G. N., Jr., « Is it possible to construct a perception machine ? », *Proceeds of the American Philosophical Society*, *134*, n° 1, 1990.

Ces articles reflètent la philosophie sous-jacente à l'un des principaux programmes de recherche en cours à l'Institut de Neurosciences. Le second est très technique et détaillé. Le troisième propose quelques réflexions sur la relation entre ce travail et d'autres dans le même domaine. Aucun article sur NOMAD – la *vraie* machine – n'a encore été publié, mais cela ne sera sans doute plus le cas lors de la publication de ce livre.

CHAPITRE 20

ADAIR, R. K., *The Great Design : Particles, Fields, and Creation*, Oxford University Press (New York), 1987.

Un très beau résumé de physique moderne avec un brin de cosmologie. Technique, mais il en vaut la peine.

ZEE, A., *Fearful Symmetry : The Search for Beauty in Modern Physics*, Macmillan (New York), 1986.

WEYL, H., *Symétrie et mathématique moderne*, Flammarion (Nouvelle Bibliothèque scientifique), 1964.

TARASOV, L., *This Amazingly Symmetrical World*, Mir (Moscou), 1986.

Trois livres sur l'importance de la symétrie en physique et ailleurs. Le livre de Zee est le plus facile. Celui de Weyl – un grand mathématicien – date d'assez tôt dans le processus.

Je ne donnerai pas ici de références concernant la mémoire en tant que principe de la nature. Les références citées précédemment devraient permettre de s'en faire une idée assez détaillée.

L'ESPRIT SANS LA BIOLOGIE : POSTFACE CRITIQUE

Ici, je suis obligé de donner une liste assez longue (et pourtant incomplète) rangée dans l'ordre des sections de la postface.

LA PHYSIQUE : L'ÉPOUVANTAIL DE SERVICE

PENROSE, R., *The Emperor's New Mind*, Oxford University Press (Oxford), 1989.

Un livre qui a beaucoup de succès, si l'on en croit le nombre d'exemplaires vendus au grand public. Contient des descriptions charmantes et lucides de physique étrange, de mesure quantique, etc. Mais la majeure partie du livre est presque totalement sans rapport avec ses objectifs et ses affirmations concernant l'esprit, comme je l'évoque dans le texte.

LOCKWOOD, M., *Mind, Brain, and the Quantum : The Compound « I »*, Basil Blackwell (Cambridge), 1989.

Les arguments d'un philosophe sur un grand nombre des questions abordées par Penrose. Peu concluant.

ZOHAR, D., *The Quantum Self : Human Nature and Consciousness Defined by the New Physics*, William Morrow (New York), 1990.

Quantique par-ci, quantique par-là, quantique partout. Un livre dont l'auteur utilise la physique comme épouvantail de service autant que faire se peut lorsqu'on est de bonne foi. Comparé au livre de Penrose, c'est néanmoins un exemple très doux du genre.

ORDINATEURS NUMÉRIQUES : LA FAUSSE ANALOGIE

HODGES, A., *Alan Turing ou l'énigme de l'intelligence*, Payot (Bibliothèque scientifique), 1988.

Biographie d'un esprit remarquable et d'un homme qui connut une fin tragique. Turing et von Neumann sont les deux théoriciens clés de l'informatique. Mais, bien sûr, beaucoup d'autres logiciens et mathématiciens avaient déjà planté le décor, comme je le fais remarquer dans le texte.

JOHNSON-LAIRD, P. N., *The Computer and the Mind*, Harvard University Press (Cambridge), 1988.

La meilleure description qui soit sur la conception de l'esprit en tant que machine.

GRAUBARD, S. R. (ed.), « Artificial intelligence », *Daedalus*, *117*, n° 1, 1988.

Série d'essais sur le thème ; certains pour, d'autres contre.

PUTNAM, H., *Représentation et Réalité*, Gallimard (NRF Essais), 1990.

Réfutation de sa propre doctrine – le fonctionnalisme des machines de Turing – par l'un des plus éminents philosophes vivant aujourd'hui.

SEARLE, J., *Minds, Brains, and Science : The 1984 Reith Lectures*, Cambridge, Harward University Press, 1984.

ANDERSON, J. A. et ROSENFELD, E. (eds), *Neurocomputing : Foundations of Research*, MIT Press (Cambridge), 1988.

ANDERSON, J. A., PELLIONISZ, A. et ROSENFELD, E. (eds), *Neurocomputing : Direction for Research*, MIT Press (Cambridge, Mass.), 1990.

Deux recueils d'articles sur divers aspects des modèles neuronaux et du connexionnisme.

QUELQUES CERCLES VICIEUX DANS LE PAYSAGE COGNITIF

GARDNER, H., *The Mind's New Science*, Basic Books (New York), 1985.

Un excellent aperçu général des sciences cognitives.

WITTGENSTEIN, L., *Investigations philosophiques*, Gallimard (Tel), 1986.

Déjà cité. Ces premières dissections des problèmes de catégorisation et de ressemblances familiales constituent un véritable travail de pionnier. L'article de G. C. Rota cité dans les références du chapitre 14 est également de mise ici.

La deuxième partie de la figure sur la catégorisation et les ensembles polymorphes (figure P-6, à droite) est extraite de Dennis *et al.*, « New problem in concept formation », *Nature*, *243*, pp. 101-102, 1973.

ROSCH, E., « Human categorization », *in Studies in Cross-Cultural Psychology*, WARREN, N. (ed.), pp. 1-49, Academic (New York), 1977.

Écrit par l'un des plus importants psychologues du domaine de la catégorisation.

BERLIN, B. et KAY, P., *Basic Color Terms : Their Universality and Evolution*, University of California Press (Berkeley), 1969.

TVERSKY, A. et KAHNEMAN, D., « Probability, representativeness, and the conjunction fallacy », *Psychological Review*, *90*, n° 4, pp.293-315, 1990.

Études de pionniers sur la catégorisation des couleurs et sur le raisonnement par analogie en général.

On trouvera les références à L. Rips et à L. Barsalou qui figurent dans le texte dans les articles de ces auteurs inclus dans *Similarity and Analogical Reasoning*, Vosniadou, S. et Ortony A. (eds), Cambridge University Press (Cambridge), 1989.

FODOR, J. A., *Representations : Philosophical Essays on the Foundations of Cognitive Science*, MIT Press (Cambridge, Mass.), 1981.

Écrit par un philosophe très prolifique, qui est aussi l'un des grands défenseurs du « mentalais ».

MARR, D., *Vision : A Computational Investigation into the Human Representation and Processing of Visual Information*, Freeman (San Francisco), 1982.

Le dernier ouvrage de cette importante figure de la psychophysique et des neurosciences, aujourd'hui décédée. Épouse la conception computationnelle mais donne un bon résumé des « premiers » processus

visuels. Les derniers chapitres, qui abordent les questions de catégorisation, sont moins bons.

MILLIKAN, R. G., *Language, Thought, and Other Biological Categories : New Foundations for Realism*, MIT Press (Cambridge, Mass.), 1984.

MILLIKAN, R. G., « Thoughts without laws ; cognitive science with content », *Philosophical Review*, XCV, n° 1, pp. 47-80, janvier 1986.

Étant l'une des figures clés de ce que j'appelle le Club des Réalistes, Millikan a émis une critique puissante et originale de ce qu'elle appelle le rationalisme du sens (équivalent, en gros, à ce contre quoi je m'élève dans la postface).

GAULD, A., « Cognitive psychology, entrapment, and the philosophy of mind », *in The case for Dualism*, Smythies, J. R. et BELOFF, J. (eds), pp. 187-253, University Press of Virginia (Charlottesville), 1989.

Il n'est pas nécessaire d'être d'accord avec le dualisme pour apprécier la violente attaque de Gauld.

SHANON, B., « Semantic representation of meaning : a critique », *Psychological Bulletin*, *104*, n° 1, pp. 70-83, 1988.

Un bon résumé des problèmes auxquels se heurtent le fonctionnalisme, l'objectivisme et la notion de représentation mentale.

PUTNAM, H., *Représentation et Réalité*, Gallimard (NRF Essais), 1990.

Déjà cité ; un morceau choisi du maître.

BRUNER, J., *Acts of Meaning*, Harvard University Press (Cambridge, Mass.), 1990.

Un superbe essai, par l'un des fondateurs des sciences cognitives modernes, qui plaide en faveur de l'importance de la narrative en tant qu'aspect de notre vie mentale.

VON HOFSTEN, C., « Catching », *in Perspectives on Perception and Action*, HEUER, H. et SANDERS, A.F. (eds), pp. 33-46, Lawrence Erlbaum Associates (Hillsdale N. J.), 1987.

Critique, venue du côté moteur, de la notion de traitement de l'information.

LANGACKER, R. W., *Foundations of Cognitive Grammar*, vol. 1 : *Theoretical Prerequisites*, Stanford University Press (Stanford), 1987.

Écrit par l'un des premiers spécialistes de cet important domaine, dont le travail a été poussé plus loin par Lakoff (voir la référence ci-dessous).

LAKOFF, G., *Women, Fire, and Dangerous Things : What Categories Reveal About the Mind*, University of Chicago Press (Chicago), 1987.

JOHNSON, M., *The Body in the Mind : The Bodily Basis of Meaning, Imagination, and Reason*, University of Chicago Press (Chicago), 1987.

Deux importants ouvrages dont les auteurs ont travaillé ensemble. Ils contiennent des références aux autres auteurs cités dans cette section. Johnson fournit les arguments qui incitent à considérer la métaphore comme un puissant produit de l'incarnation. Le texte de Lakoff, plus détaillé, comprend une histoire des idées, une critique de la notion

d'esprit en l'absence de la biologie, et une plaidoirie en faveur de la reconnaissance de l'incarnation en tant que base de la signification et de l'esprit. Il sert de base à une grande part de la dernière partie de la postface. Par ailleurs, Lakoff y décrit longuement la grammaire cognitive que j'ai résumée ici. Le présent ouvrage et ma trilogie peuvent être considérés comme une réponse théorique à la question : « Comment l'esprit s'incarne-t-il ? » Cette question est soulevée par les travaux cités ci-dessus de telle sorte qu'elle ne peut plus être ignorée.

LE LANGAGE : POURQUOI L'APPROCHE FORMELLE ÉCHOUE

CHOMSKY, N., *Cartesian Linguistics*, Harper & Row (New York), 1966.

CHOMSKY, N., *Rules and Representations*, Columbia University Press (New York), 1980.

Deux livres écrits par le plus important linguiste des dernières années, défenseur de l'approche formelle et le plus puissant de ses adeptes.

LIGHTFOOT, D., *The Language Lottery : Toward a Biology of Grammars*, MIT Press (Cambridge, Mass.), 1982.

Compte rendu plein d'informations, par l'un des émules.

JACKENDOFF, R., *Consciousness and the Computational Mind*, MIT Press (Cambridge, Mass.), 1987.

Le meilleur résumé qui soit alliant la conception du langage en tant que syntaxe à celle de l'esprit en tant que machine. Résultat : la conscience en tant qu'épiphénomène. Évidemment, je récuse ce point de vue.

BRESNAN, J. (ed.), *The Mental Representation of Grammatical Relations*, MIT Press (Cambridge, Mass.), 1982.

Description détaillée de la grammaire fonctionnelle lexicale ; l'une des bases de l'analyse de Pinker sur l'acquisition du langage. Difficile à lire pour les néophytes.

PINKER, S., *Language Learnability and Language Development*, Harvard University Press (Cambridge, Mass.), 1984.

Une analyse intelligente et pénétrante. Technique.

DONALDSON, M., *Children's Minds*, Norton (New York), 1978.

Moins technique. Charmant et néanmoins presque mortel dans ses attaques contre l'idée de l'existence d'un dispositif d'acquisition du langage.

LEVELT, W. J. M., *Speaking : From Intention to Articulation*, MIT Press (Cambridge, Mass.), 1989.

GRICE, H. P., « Logic and conversation », *in Studies in Syntax*, vol. 3, pp. 41-58, COLE, P. et MORGAN, J. L. (eds), Academic (New York), 1967.

Levelt aborde du point de vue concret, et de façon détaillée, le problème

de l'expression orale. Cela ne l'empêche pas de se placer dans le cadre des sciences cognitives conventionnelles. Ce volume est probablement le meilleur qui existe sur le thème. Grice analyse de façon extrêmement originale les conditions nécessaires à l'établissement d'échanges oraux efficaces.

VIGOTSKI, L. S., *Pensée et langage*, Messidor-Editions sociales (Terrains), 1985.

Ce penseur soviétique (un collègue de A. Luria) a mis l'accent sur les échanges interpersonnels et sociaux, et sur l'« intériorisation » du discours au service de la pensée.

PERCY, W., *The Message in the Bottle : How Queer Man is, How Queer Language is, and What One has to do with the Other*, Farrar, Straus & Giroux (New York), 1976.

À la fois amateur, émouvant et profond. La preuve qu'on n'a pas besoin de beaucoup de diplômes de haut niveau pour réussir à penser, et que le fait d'être médecin n'est pas toujours rédhibitoire. Percy était un bon romancier mineur.

KLIMA, E. et BELLUGI, U., *The Signs of Language*, Harvard University Press (Cambridge, Mass.), 1978.

Une analyse des travaux de pionnier d'Ursula Bellugi montrant que le langage des sourds-muets possède une syntaxe, des dialectes et d'autres caractéristiques du langage parlé.

BICKERTON, D., *Language and Species*, University of Chicago Press (Chicago), 1990.

Une courageuse tentative de rendre compte de l'évolution du langage parlé à travers un « pidgin » intermédiaire ou protolangue. Provocateur.

LIEBERMAN, P., *Uniquely Human : The Evolution of Speech, Thought, and Selfless Behavior*, Harvard University Press (Cambridge, Mass.), 1991.

Une bonne description, destinée à un public élargi, écrite par l'un des grands spécialistes de l'évolution de l'appareil vocal.

KELLER, H., *The Story of My Life*, 1902. Réédition, Doubleday (New York), 1954.

Poignant. Très à propos pour achever cette liste, puisqu'il décrit l'acquisition de compétences linguistiques contre toute attente.

Sources des citations et des illustrations

CITATIONS

Frontispice

Empédocle, *in* Voilquin, J. (trad.), *Les penseurs grecs avant Socrate – de Thalès de Milet à Prodicos*, Garnier-Flammarion, 1964, p. 132.
Feynman, R., *The Character of Physical Law*, MIT Press (Cambridge, Mass.), 1965, p. 125.

Dédicace

L'Ecclésiaste, 1 : 18, vers 250 av. J.-C.

Chapitre 1

Descartes, R., *Le Discours de la méthode* [1637], Flammarion, 1966.
De Unamuno, M., *Del sentimiento trágico de la vida* [1912]. Réédition Editorial Losada (Buenos Aires), 1964.

Chapitre 2

Whitehead, A. N., *Science and the Modern World*, Macmillan (New York), 1925, pp. 2-3

Chapitre 3

Maxwell, J. C., *in* Barrow, J. D. et Tipler, F. J., *The Anthropic Cosmological Principle*, Oxford University Press (Oxford), 1986, p. 545.
Diderot, D., *Le Rêve de d'Alembert* [1830], éditions Marcel Didier.

Chapitre 4

Adams, H., *The Education of Henry Adams*, Houghton Mifflin (New York), 1961.

Chapitre 5

Darwin, C., *in* Zinsser, H., *As I Remember Him : The Biography of R.S.* [1939], Peter Smith (Magnolia, Mass.), 1970.

Darwin, C., *in* Barrett, P. H., Gautrey, P. J., Herbert, S., Kohn, D. et Smith, S. (eds), *Charles Darwin's Notebooks, 1836-1844 : Geology, Transmutation of Species, Metaphysical Enquiries*, Cornell University Press (Ithaca, N. Y.), 1987, p. 539.

Chapitre 6

Spitzer, N., « The chicken and the egg, together at last » (compte rendu de *Topobiology : An Introduction to Molecular Embryology*), *The New York Times Book Review*, 22 janvier 1989, p. 12.

Chapitre 7

Pascal, B., *Pensées* [1670], LGF (Le Livre de poche-Classique), 1962.

Chapitre 8

Anonyme.

Chapitre 9

Crick, F. H. C., « Neural edelmanism », *Trends in Neuroscience*, *12*, n° 7, juillet 1989, p. 247.

Chapitre 10

Wittgenstein, L., *in* Pitcher, G. (ed.), *Wittgenstein : The Philosophical Investigations*, Macmillan (Londres), 1968, p. 465.

Voltaire, *Dictionnaire philosophique*, Flammarion, 1964.

Chapitre 11

James, W., *Psychology : Briefer Course*, Harvard University Press (Cambridge), 1984, p. 401.

Chapitre 12

Focillon, H., *Vie des formes ; Éloge de la main* (4e éd.), PUF (Quadrige), 1990.

Chapitre 13

Freud, S., *in* Strachey, J. (trad. et ed.), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol. 14, Hogarth Press (Londres), 1976, p. 280.

Valéry, P., « Mauvaises pensées », *in Œuvres*, Tome II, Gallimard, 1960. Réédition 1988 (Bibliothèque de la Pléiade).

Chapitre 14

Bridgman, P. W., « Quo Vadis », *Daedalus*, hiver 1958, p. 93.

Chapitre 15

Holmes, O. W., *The Complete Works of Oliver Wendell Holmes*, Scholarly Press (St Clair Shores, Mich.), 1972.

Einstein, A., *in* Atchity, K., *A Writer's Time*, Norton (New York), 1986, p. 180.

Chapitre 16

Planck, M., *in* Barrow, J. D. et Tipler, F. J., *The Anthropic Cosmological Principle*, Oxford University Press (Oxford), 1986, p. 123.
Allen, W. Reproduit avec la permission de Monsieur Allen.

Chapitre 17

Schopenhauer, A., *Counsels and Maxims*, Scholarly Press (St Clair Shores, Mich.), 1981.

Chapitre 18

Freud, S., *in* Bonaparte, M., Freud, A. et Kris, E., *The Origins of Psycho-Analysis*, Basic Books (New York), 1954, p. 120.

Chapitre 19

La Mettrie, J. O. de, *L'homme machine*, Denoël (Paris), 1981.

Chapitre 20

Einstein, A., *in* March R. H., *Physics for Poets*, Contemporary Books (Chicago), 1978, p. 135.
Valéry, P., *Monsieur Teste,* Gallimard (Paris), 1978.

Postface (citation dans le texte)

Quine, W. V., *The Ways of Paradox*, Random House (New York) 1966, p. 66.

*
* *

ILLUSTRATIONS

Figures 1-1, 1-2, 4-1 et 5-1 : Mary Evans Picture Library, Londres.

Figures 2-1 et 20-1 : Copyright © American Institute of Physics. Reproduit avec la permission des auteurs.

Figure 2-2 : Barrow, H., Blakemore, C. et Weston-Smith, M. (eds.), *Images and Understanding*, Cambridge University Press (New York), 1990, p. 261. Copyright © 1990 Cambridge University Press. Reproduit avec la permission des auteurs.

Figure 3-1 : Blakemore, C., Greenfield, S. (eds.), *Mindwaves*, Basil Blackwell (New York), 1987, p. 4. Copyright © 1987 Basic Blackwell. Reproduit avec la permission des auteurs.

Figures 3-2 et 3-3 : Edelman, G. M., *Topobiology : An Introduction to Molecular Embryology*. Copyright © 1988 Basic Books, Inc. Reproduit avec la permission de HarperCollins Publishers.

Figure 3-5 (en haut à gauche) : Pearson, K. G. et Goodman, C. S., « Correlation of variability in structure with variability in synaptic connection of an iden-

tified interneuron in locusts », *Journal of Comparative Neurology*, *184*, 141-165, 1979. Reproduit avec la permission de Corey S. Goodman.

Figure 3-5 (en bas) : Merzenich, M. *et al.*, « Topographic reorganization of somatosensory cortical areas 3b and 1 in adult monkeys following restricted deafferentation » et « Progression of change following median nerve section in the cortical representation of the hand in areas 3b and 1 in adult owl and squirrel monkeys », *Neuroscience*, *8*, 33-55, 1983 et *Neuroscience*, *10*, 639-665, 1983. Reproduit avec la permission de Michael M. Merzenich.

Figure 3-5 (en haut à droite) : Macagno, E. R., Lopresti, V. et Levinthal, C., « Structure and development of neuronal connections in isogenic organisms : variations and similarities in the optic systems of *Daphnia magna* », *Proceedings of the National Academy of Science*, *70*, 57-61, 1973. Reproduit avec la permission d'Eduardo R. Macagno.

Figure 4-1 (à gauche) : Durant, W. et A., *The Story of Civilization : Rousseau and the Revolution*, vol. x, Simon & Schuster (New York), 1967. Reproduit avec la permission des exécuteurs de la succession d'Ethel Durant.

Figure 4-2 : Kanizsa, G., *Organization in Vision : Essays on Gestalt Perception*, Praeger (New York), 1979, pp. 78 et 74. Copyright © 1979 Gaetano Kanizsa. Reproduit avec la permission de Greenwood Publishing Group, Inc., Westport, Conn.

Figure 5-3 : d'après Lewontin, R., *The Genetic Basis of Evolutionary Change*, Columbia University Press (New York), 1974. Copyright © 1974 Richard Lewontin. Reproduit avec la permission des auteurs.

Figure 5-4 (à gauche) : Iltis, H., *The Life of Mendel* (Hafner Publishing Co. (New York), 1966. Reproduit par Unwin Hyman, de Harper Collins Publishers, Ltd.

Figure 5-5 : Barrett, P. H., Gautrey, P. J., Herbert, S., Kohn, D. et Smith, S. (transc. et eds.), *Charles Darwin's Notebooks, 1836-1844 : Geology, Transmutation of Species, Metaphysical Enquiries.* Initialement publié par le British Museum (Histoire Naturelle). Copyright © 1987 Paul H. Barrett, Peter Gautry, Sandra Herbert, David Kohn, Sydney Smith. Reproduit avec la permission de l'éditeur, Cornell University Press.

Figure 5-6 : Stebbins, G. L., *Darwin to DNA, Molecules to Humanity*, W. H. Freeman and Company (New York) 1982. Copyright © 1982 W. H. Freeman and Company. Reproduit avec la permission des auteurs.

Figure 6-2 : Anderson, C. M., Zucker, F. H. et Steitz T. A., « Space-filling models of kinase clefts and conformation changes », *Science*, *204*, 375-380, 1979. La figure se trouve p. 376. Copyright © 1979 American Association for the Advancement of Science. Reproduit avec la permission des auteurs.

Figure 6-3 (en haut) : Edelman, G. M., « Cell adhesion molecules : a molecular basis for animal form », *Scientific American*, *250*, n° 4, 118-129, 1984. La figure se trouve pp. 120-121. Copyright © 1984 Scientific American, Inc. Tous droits réservés. Reproduit avec la permission des auteurs.

Figure 6-3 (en bas) : Romer, A., *The Vertebrate Body* (5e éd.), Saunders College Publishing (Orlando, Floride), 1963, p. 119. Copyright © 1977 Saunders College Publishing. Reproduit avec la permission de la maison d'éditions.

Figure 6-5 : reproduite avec la permission de Walter J. Gehring, université de Bâle, Suisse.

Figures 9-1, 9-2, 9-4, 9-5, 11-1, 12-4 et 12-5 : Edelman, G. M., *The Remembered Present : A Biological Theory of Consciousness.* Copyright © 1989 Basic Books. Reproduit avec la permission de Harper Collins Publishers.

Figure 9-2 : reproduite avec la permission de Semir Zeki, University College, Londres.

Figure 9-3 (en haut) : Edelman, G. M., *Neural Darwinism : The Theory of Neuronal Group Selection.* Copyright © 1987 Basic Books, Inc. Reproduit avec la permission de Harper Collins Publishers.

Figure 9-6 (en haut et en bas à gauche) : Edelman, G. M., Sporns, O., Reeke, G. N., Jr., « Synthetic neural modeling : comparisons of population and connectionist approaches », *in* Pfeifer, R., Schreter, Z., Fogelman-Soulie, F. et Steels, L. (eds), *Connectionism in Perspective,* Elsevier Science Publishers (New York), 1989, pp. 120 et 126. Copyright © 1989 Elsevier Science Publishers. Reproduit avec la permission des auteurs.

Figure 9-6 (en bas à droite) : Edelman, G. M., *et al.*, « Synthetic neural modeling », *in* Edelman, G. M., Gall, W. E. et Cowan, W. M. (eds), *Signal and Sense*, The Neuroscience Research Foundation (New York), 1990, p. 698. Reproduit avec la permission des auteurs.

Figure 12-1 : Lieberman, P., *The Biology and Evolution of Language*, Harvard University Press (Cambridge), 1984. Copyright © 1984 Le président et les membres de Harvard College. Reproduit avec la permission de l'auteur.

Figure 12-2 (en haut) : Penfield, W. et Roberts, L., *Speech and Brain Mechanisms.* Copyright © 1959 Princeton University Press, renouvelé 1987. Figure X-4, p. 201, reproduite avec la permission des exécuteurs littéraires de Penfield Papers et de Princeton University Press.

Figure 12-2 (en bas) : Schiller, F., *Paul Broca : explorateur du cerveau,* éditions Odile Jacob, 1990. Dr. Juster, Institut de parasitologie, École Pratique, Paris. Reproduit avec la permission de l'auteur.

Figure 13-1 : copyright 1983 Sigmund Freud Copyrights. Reproduit avec la permission de A. W. Freud *et al.*, avec l'autorisation de Mark Patterson & Associates.

Figure P-3 (en haut) : reproduite avec la permission du Computer Museum, Boston.

Figure P-3 (en bas) : reproduite avec la permission de N-Cube, Belmont, Calif.

Figure P-6 (à droite) : Dennis, I., Hampton, J. A. et Lea, S. E. G., « New problem in concept formation », *Nature, 243*, 101, 1973. Copyright © 1973 Macmillan Magazines, Ltd. Reproduit avec la permission de Ian Dennis et *Nature.*

Index thématique

Acides aminés, 76, 79.
Activité automatique (interrompue par des nouveautés), 188.
Adhérence cellulaire, 84-87.
ADN, 75-79 ; réplication, 270.
Aire visuelle, 115.
Altruisme, 66-68.
Anosognosie, 245.
Apprentissage, 45-46, 128-131.
Arborisation, 42-44.
ARN, 79 ; réplication, 270.
Attention, 186-189.
Augustin (saint), 220, 221.
Auto-catégorisation, 158.
Avery, Oswald, 78.
Axiomatiques, 200-201.

Barlow, Horace, 125.
Bartlett, Sir Frederic, 56.
Bases nucléotidiques, 75-76.
Behaviorisme, 25-26, 55-57.
Berkeley, George, 53, 280.
Biologie, 314-317 ; liens avec la psychologie, 52.
Bisiach, Eduardo, 245.
Blastoderme, 82.
Bohr, Niels, 263, 285.
Brentano, Franz, 18.
Broca, Paul, 56, 169, 172.
Brouwer, L.E.J., 221.
Bruner, Jérome, 229, 302.
Burnet, Sir Frank MacFarlane, 102.

Capacité crânienne (évolution) 69-70.
Cartes, 96 ; anatomiques, 46 ; connectées par des fibres réentrantes, 115-117 ; durant le développement, 87-89 ; fluctuations des frontières, 45-46 ; formation, 38-42 ; locales, 119-121 ; connexions parallèles et réciproques, 113-114 ; aire visuelle, 115.
Cartographies globales, 117-122, 166 ; altération de l'attention et sélection, 189 ; catégorisation, 195-196 ; concept, 143 ; Darwin III, 123-124 ; classes, 143.
Catégories radiales, 328-329.
Catégorisation, 307-312 ; maladies mentales, 238-240 ; cartographies globales, 195-196 ; ensembles polymorphes, 309-310 ; sélection, 222-223.
Causalité, 205.
Cellules cérébrales, 32.
Cellules nerveuses, 36-37 ; voisinage 89-90.
Cerveau, anatomie, 38-40 ; ordinateur, 44, 300-301 ; fonctionnalisme, 291-294 ; réseau, 42 ; noyaux, 33 ; organisation, 37-38 ; taille, complexité et comportement, 70-72 ; structures, 18-20 ; machine de Turing, 299-300.
Cervelet, mémoire, 138-139.
Chomsky, Noam, 318-320.
Church (thèse de), 291, 306.
Cognitivisme, 29-30.
Colonnes d'orientation (syntonisation), 125.
Conception computationnelle, 307.
Connaissance, hétérogénité, 232 ; boucles, 194.

Connexions, cortex, 32-34 ; neurones, 34-36.
Conscience, 55, 93-94, 147-163 ; avantage évolutif, 176-177 ; définition, 147-148, 220-221 ; évolution, 176 ; température, 207 ; médication, 161 ; modèles, 153 ; mystique, 183 ; propriétés, 181-182 ; sensation, 149-151 ; sélection naturelle, 195-196 ; explication scientifique, 183-184 ; hypothèses théoriques, 149-150 ; théorie, 147-148.
Conscience primaire, 148, 154-163 ; mémoire conceptuelle, 157 ; tableau, 158 ; évolution, 195-197 ; développement, 155-156 ; propriétés jamesiennes, 160-197 ; évolution des fonctions, 155-156 ; système nerveux, 155 ; hypothèse des sensations, 149-151 ; réentrée, 157 ; sélection de groupes, 160 ; conscience d'ordre supérieur, 161 ; hypothèses théoriques, 178-179.
Cortex cérébral, 31-32.
Cortex encorhinal, 140.
Cortex frontal, 143-144.
Cortex visuel, 125-126.
Cosmologie, 261-262, 268.
Craik, K.J.W., 303.
Crick, Francis, 109, 125, 127, 128.

Danto, Arthur, 205-206.
Darwin, Charles, 61-63 ; 68-69 ; programme, 61-68.
Darwin III, 119-124 ; comportement, 254-255 ; carte globale, 122 ; mémoire, 135-136 ; sortie par cartes réentrantes, 122 ; signal de valeurs, 254.
Déictique, 319.
Dendrite, 42.
Descartes, René, 15, 23, 52-53, 192.
Deux-icité, 221.
Déterminisme, 223-224.
Développement, évolution, 71-72 ; dépendances spatiales, 38-39 ; études, 58-60, 96.
Diderot, Denis, 36, 219.
Dilthey, Wilhelm, 232.
Dualisme des substances, 25.
Dualistes des propriétés, 26.
Dynamique des populations malthusiennes, 125.

Ebbinghaus, Hermann, 55.
Einstein, Albert, 23-25, 207.
Elliot, Stuart, 78.
Embryon, 81-87.
Émotions, 231.
ENIAC, 292.
Entropie, 270.
Épigénétique, événements, 39-42 ; interactions entre espèces, 66-68.
Épistémologie, 209.
Espace, 267.
Esprit, 15-21 ; analogie avec l'ordinateur, 200 ; bases biologiques, 226 ; concept de, 17 ; relation avec des phénomènes nerveux, 19 ; doctrine, 16, 25, 27 ; mortalité, 226 ; sous-bassement matériel, 31 ; comme processus, 18-20 ; replacé dans la nature, 25 ; ce qui est facile et ce qui est difficile à comprendre, 185.
Éthologie, 66-67.
Êtres représentationnels, 205.
Évolution et développement 71-72 ; sélection naturelle, reconnaissance, 105.
Évolutionniste, hypothèse, 150.

Faraday, Michael, 27.
Fermi, Enrico, 106.
Fodor, Jerry, 301.
Fonctionnalisme, 292-300.
Fonctions cérébrales supérieures, 175.
Fonctions mentales, 27.
Forme, phénomènes de, 56.
Freud, Sigmund, 190-191, 236, 240.
Fritsch, Gustav, 56.

Galilée, 23.
Galois, Evariste, 264-265.
Ganglions de la base, dans l'attention, 188 ; mémoire, 138-140.
Gedankenesperiment, 149.
Générateur de nombres pseudo-aléatoires, 252.

Gènes, changements dans la fréquence des, 63-64 ; modifications morphologiques, 68 ; homéotiques, 88 ; expression séquentielle, 85.
Gödel, Kurt, 199-200.
Grammaire, générative, 320-323 ; cognitive, 321-323, 326-330 ; lexique, 171-172.

Hebb, Donald, 57.
Hegel, G.N.F., 232.
Heisenberg, principe d'incertitude de, 24, 284.
Hering, Ewald, 55.
Hilbert, David, 200.
Hippocampe, mémoire, 138-140.
Hitzig, Julius, 56.
Homéostats, 124.
Homéotiques, gènes, 87.
Hominidés, comportement des, 70.
Homoncules, 108.
Hull, Clark Leonard, 56.
Hume, David, 53, 218.
Hundert, Edward, 240.
Husserl, Edmund, 207.

Idéalisme, 212-213.
Incarnation conceptuelle, 324.
Inconscient, 190-192 ; théories psychologiques freudiennes, 191 ; refoulement, 191.
Intelligence artificielle, 298-300.
Intentionnalité, 17-18, 95, 148, 303 ; maladie neuroligique, 246.

James, William, 19, 55.
Johnson, Mark, 315, 325.

Kanizsa, Gaetano, 199.
Kant, Emmanuel, 53-54, 232.
Koffka, Kurt, 56.
Köhler, Wolfgang, 56.
Kreisler, Fritz, 137.

Lakoff, George, 171, 314-315, 323-330.
Langacker, Ronald, 171, 323-328.
Langage, acquisition, 317-323 ; grammaire cognitive, 326; définition, 320 ; évolution, 165-166 ; conception formelle, 316-324 ; grammaire, 317 ; conscience d'ordre supérieur, 165, 324 ; mémoire, 312-314 ; forme et signification, 317 ; relation avec la pensée, 317-318 ; âme, 59 ; syntaxe et sémantique, 315-316.
Langer, Suzanne, 231.
Lashley, Karl, 57.
Libre arbitre, 223.
Lie, Sophus, 265.
Limbique, système, 156-157.
Lipmann, Fritz, 263.
Locke, John, 53.
Lois de conservation, et symétrie, 265-268.
Lorenz, Konrad, 59.
Lumières, philosophie et Siècle des, 217-219.
Lymphocytes, 101-105.

Maladie mentale, 235-248 ; altération des cartes réentrantes et catégorisation, 237-238 ; anosognosie, 245-246 ; dissociation, 241 ; troubles fonctionnels, 236 ; lésions sous-jacentes, 242 ; troubles neurologiques, 235, 242 ; développement psychologique, 236 ; réintégration, 247 ; schizophrénie, 242-244 ; théorie de la sélection des groupes neuronaux, 236-238 ; mémoire des valeurs-catégories, 242.
Marr, David, 303.
Mayr, Ernst, 99.
Mécanique quantique, 23, 263.
Medawar, Sir Peter, 27.
Mémoire, 56-57, 134-142, 221-222 ; ganglions de la base, 138-140 ; cervelet, 138 ; cortex, 139 ; appendices corticaux, 139 ; Darwin III, 136-137 ; développement, 273-274 ; processus héréditaire, 272-274 ; hippocampe, 139 ; langage, 312-314 ; sélection de groupes neuronaux, 135 ; changements successifs, 137 ; recatégorisation, 135, 137 ; symbolique, 166 ; symétrie, 270-275 ; élément temporel, 218 ; théorie de la sélection des

groupes neuronaux, 135-136 ; types, 271.
Mendel, Gregor, 65.
Mentalais, 301-302.
Merzenich, Michael, 127.
Mesure quantique, 282-284.
Métaphore, 311, 324-325.
Métonymie, 311, 324-325.
Michod, Richard, 127.
Millikan, Ruth Garret, 316.
Mise en correspondance sémantique, 168-169, 171-172, 197.
Modell, Arnold, 240.
Monod, Jacques, 191.
Morale, 225-226.
Morgan, C. Lloyd, 57.
Morphologie, 69-70, 47.

N-CUBE, 292.
Neuromodulateur, 38.
Neurones, 31-36 ; connexions, 34-36 ; nombre de, 32 ; sensibilité aux stimulations, 38 ; formes, 34-36.
Neurophysiologie, 57.
Névroses, 236.
Noétiques, engins, 254.
NOMAD, 254-255.
Nombres, concept de nombre et temps, 220-221.
Noyaux, 33-34.

Objectivisme, 29, 93-94, 303-307 ; définition, 304.
Objet conscient artificiel, 249-259 ; capacité future à fabriquer un, 256-258 ; NOMAD, 255.
Objets, classement, 45.
Onsager, Lars, 245.
Ordinateur, 288-300 ; algorithmes, 289-292 ; intelligence artificielle, 299-300 ; symboles des objets physiques, 292-294 ; machine de Turing, 289-291 ; simulation, 250-251.
Organisation biologique, niveaux d', 194.

Panpsychisme, 280-282.
Paradigme stimulus-réponse, 56.
Pavlov, Ivan, 55.
Penrose, Roger, 286.
Pensée, 51, 228-230 ; expériences de, 149 ; corrélations entre le neuronal et la signification de la, 231 ; explication neuroscientifique, 230 ; pensée pure, 228 ; relation avec le langage, 318-320.
Percy, Walker, 322-323.
Phénotype, 61, 63, 66, 75, 78, 79, 115, 199, 330.
Philosophie, 205-216 ; relation avec la science, 205-207 ; identité, 216.
Physique, 280-288 ; révolution, 282 ; échelles naturelles, 283.
Piaget, Jean, 57.
Planck, Max, 91, 263, 282.
Platon, 314.
Polypeptide, 76.
Populations, penser en termes de, 63, 99-100, 315.
Post-synaptique, neurone, 34.
Principe de complémentarité, 285.
Programme moteur, 138, 188.
Prolongements cellulaires, 112.
Propriétés sensorielles, niveaux, 198.
Prosopagnosie, 162, 183, 241.
Protéines, repliement et fonction, 76-79.
Psychologie, branche de la biologie, 316 ; liée à la biologie, 52 ; écoles de pensée, 53 ; psychologie sociale, 59.
Purves, Dale, 127.
Putnam, Hilary, 294-296, 312.

Quine, W. V., 307.

Rabi, Isidor, 286.
Rachmaninoff, Sergei, 137.
Ramón y Cajal, Santiago, 57.
Rationalisme cartésien, 51- 53.
Recatégorisation, 135, 137.
Réalisme pondéré, 210-211.
Reconnaissance, 98-108 ; diversité anatomique, 110 ; cerveau, 107-108 ; définition, 100-101 ; évolution, 106-107 ; système immunitaire, 103-105 ; mécanismes, 128.
Réductionnisme, 217-218, 223-226.
Refoulement, 191.

Réintégration, maladie mentale, 247.
Relation de transfert, 238.
Relativité, théorie, 263-264.
Répertoire primaire, 116, 128.
Répertoire secondaire, 127.
Représentation mentale, 300-302, 308-311.
Réseaux neuronaux, architecture, 299.
Romanes, George, 57.
Rosch, Eleanor, 310.
Rous, Peyton, 78.
Russell, Bertrand, 306.

Schacter, Daniel, 241.
Schizophrénie, 242-245.
Science, en relation avec la philosophie, 205-207.
Sciences cognitives, 27.
Searle, John, 295-296, 314.
Sélection clonale, 104.
Sélection des groupes, 68.
Sélection naturelle, 61-63, 195-196 ; conscience, 196-197 ; programme darwiniste, 63-67 ; reproduction différentiée, 88.
Sélection somatique, 194-195.
Sélectionniste, théorie, 213.
Sentiments, 231.
Sherrington, Sir Charles, 57.
Signaux nerveux, 37.
Signaux réentrants, 114, 117.
Signification, 224, 229, 294-296, 301 ; représentation sémantique, 303, symboles, 298.
Skinner, B. F., 56.
Sociobiologie, 67-68.
Spencer, Herbert, 58.
Split-brain, cas de, 247.
Structures neuronales, 132, 256, 299.
Symboles, signification, 298.
Symétrie, 262-270 ; rupture, 268, 274 ; lois de conservation, 265-268 ; globale, 266 ; locale, 266 ; mémoire, 266-275 ; types, 262.
Synapse, 32-35 ; silencieuses, 44.
Système nerveux, anatomie, 40 ; réponses, 298.
Système réplicatif, structure apériodique, 272.
Systèmes sélectifs, catégorisation, 220 ; simulation, 251.
Systèmes connexionistes, 298-300.
Systèmes thalamo-corticaux, 155-156.
Szilard, Leo, 106, 107.

Temps, 267, connexion avec l'idée de nombre, 299.
Téléonomie, 210-211.
Théorie de la relativité générale, 263.
Théorie de grande unification, 267.
Théorie de l'instruction, 102.
Théorie de la gravitation quantique, 284.
Thermodynamique, deuxième principe, 269.
Thorndike, Edward, 56.
Tinbergen, Nikolaas, 59.
Titchener, Edward, 55.
Toit optique, 38-42.
Topobiologie, 75-90 ; rôles des CAM et des SAM, 87 ; division cellulaire, 81-86 ; forme caractéristique, 84 ; diversité, 89 ; gènes, 85.
Transducteurs sensoriels, 34.
Tronc cérébral, 138-140.
Tube neural, 84.
Turing, Alan, 289.

Uhlenbeck, George, 263.

Vaucanson, Jacques de, 249.
Vérité, 205.
Vision scientifique, hypothèses, 207.
Voies réentrantes, altérations dans les maladies mentales, 237, 242.
Von Helmholtz, Hermann, 55.
von Neumann, John, 283.

Wallace, Alfred, 61.
Warhol, Andy, 258.
Watson, John, 56.
Wernicke, 169 ; aire de, 172.
Wertheimer, Max, 56.
Whitehead, Alfred North, 23, 306.
Wigner, Eugène, 283.
Wittgenstein, Ludwig, 308, 309.
Wundt, Wilhelm, 55

Remerciements

Je suis entièrement responsable des opinions exprimées dans ce livre, ainsi que de toutes les erreurs résiduelles qu'il est susceptible de contenir. Mais sans les nombreuses personnes qui m'ont fait part de leurs réactions et de leurs critiques, le fardeau aurait été bien plus lourd. Je tiens à exprimer ma gratitude envers tous ceux qui ont fait preuve de générosité à mon égard, sachant que je ne pourrai pas mentionner ici les noms de toutes les personnes qui m'ont aidé.

Susan Hassler, éditrice de l'Institut de Neurosciences, a mis son expertise à contribution à tous les stades. Son aide a été inestimable. Kathryn Crossin, qui avait déjà participé à la révision de ma trilogie, a aussi émis un certain nombre d'importantes remarques d'ordre éditorial en ce qui concerne le présent ouvrage. La même chose vaut pour George N. Reeke, Jr., mon proche collaborateur, dont les contributions à la conception d'automates de reconnaissance sont fondamentales. Olaf Sporns, l'un de mes anciens étudiants et aujourd'hui mon collègue, a émis de nombreuses suggestions pleines d'imagination et mis généreusement ses talents de dessinateur à ma disposition pour la confection d'un grand nombre des figures du livre. Ses contributions ont été essentielles.

Je suis particulièrement reconnaissant à Susan Borden et Henry G. Walter, Jr. qui, chacun de leur côté, m'ont fait des suggestions tout à fait pertinentes concernant l'organisation du livre. W. Einar Gall, directeur scientifique de l'Institut, ainsi que Giulio Tononi et Joseph Gally (tous deux membres de l'Institut), ont émis des critiques très importantes – de même que Detlev Ploog, durant son séjour à l'Institut en tant que membre invité.

Ce livre est le seul de mes livres que ma femme, Maxine, ait lu dès les stades les plus précoces. Ses réactions positives m'ont aidé à donner forme à certains des points de vue exprimés dans les dernières parties. Je l'en remercie, et également pour m'avoir encouragé en ce qui concerne cette question de l'esprit et d'autres.

Enfin, j'ai placé en exergue dans la traduction française une phrase qui a servi de titre à la version originale de ce livre. C'est un hommage à la découverte de l'esprit par les Grecs au VIe siècle avant J.-C. Elle est extraite d'un fragment d'Empédocle, physicien, poète, l'un des premiers philosophes à soutenir une conception matérialiste de l'Esprit. Son idée de la perception, résultat de l'adéquation d'entités matérielles aux pores de notre corps, convient plus à l'odorat qu'à la vision. Mais son cœur (lieu de l'Esprit, selon lui) allait dans le bon sens.

Liste des illustrations

1-1. René Descartes (1596-1650).
1-2. William James (1842-1910).
2-1. Galileo Galilei (1564-1642).
2-2. Diagramme du système visuel extrait du *Traité de l'homme* de Descartes.
3-1. La surface du cortex cérébral humain.
3-2. Plusieurs organisations de la matière de l'esprit.
3-3. Le développement du cerveau.
3-4. Cartographie de l'œil et de ses champs visuels sur le cerveau.
3-5. Variabilité des structures neuronales.
4-1. Emmanuel Kant (1724-1804).
4-2. Phénomènes de forme.
5-1. Charles Darwin (1809-1882).
5-2. Penser en termes de populations.
5-3. Les modifications de la fréquence des gènes peuvent être reliées au processus de sélection naturelle lui-même.
5-4. La synthèse moderne.
5-5. Extrait des cahiers de Darwin.
5-6. L'augmentation remarquable de la capacité crânienne au cours de deux millions d'années d'évolution de l'espèce humaine.
6-1. Du décodage de l'information génétique à la fabrication des protéines.
6-2. Repliement et fonction des protéines.
6-3. Le début du développement de l'embryon de poulet.
6-4. L'adhérence cellulaire.
6-5. Un exemple de topobiologie aberrante.
8-1. Le système immunitaire fonctionne comme un système sélectif de reconnaissance.
8-2. La suite sans fin des homoncules emboîtés.
9-1. Une théorie sélectionniste de la fonction cérébrale.
9-2. Les multiples cartes des aires visuelles du cerveau.
9-3. Les groupes neuronaux.
9-4. La réentrée.
9-5. Une cartographie globale.
9-6. Darwin III, un automate de reconnaissance qui se comporte comme une cartographie globale.

10-1. Deux conceptions de la mémoire.
10-2. Les appendices corticaux.
11-1. Un modèle de la conscience primaire.
12-1. Le tractus supralaryngé chez l'être humain.
12-2. Aires du cerveau servant à la production du langage parlé.
12-3. Bootstrapping sémantique.
12-4. Comment apparaît la conscience d'ordre supérieur.
12-5. L'évolution de la conscience dépend de l'apparition de nouvelles morphologies.
13-1. Sigmund Freud (1856-1939).
14-1. Les niveaux de l'organisation biologique et les boucles de la connaissance.
18-1. Maladies du système nerveux et maladies de l'esprit.
19-1. Jacques de Vaucanson (1709-1782).
19-2. NOMAD (engin à structure neuronale et à adaptabilité multiple).
20-1. Divers types de symétries.
20-2. Divers types de mémoires.
P-1. Les échelles de grandeurs naturelles définies par la physique.
P-2. Une machine de Turing.
P-3. Deux ordinateurs réels.
P-4. Algorithme permettant de faire cuire un œuf.
P-5. Quelques aspects de l'objectivisme et du fonctionnalisme.
P-6. Catégorisation et ensembles polymorphes.
P-7. Arbre typique d'une grammaire générative.
P-8. Exemple des processus d'une grammaire cognitive.

Table des matières

Préface 11

PREMIÈRE PARTIE

Les problèmes

Chapitre 1 : L'esprit 15

Chapitre 2 : Replacer l'esprit dans la nature 23

Chapitre 3 : La matière de l'esprit 31

DEUXIÈME PARTIE

Les origines

Chapitre 4 : Les bases biologiques de la psychologie 51

Chapitre 5 : L'esprit et sa morphologie, ou comment achever le programme de Darwin 61

Chapitre 6 : La topobiologie : leçons d'embryologie 75

Chapitre 7 : Les problèmes revus et corrigés 91

TROISIÈME PARTIE

Propositions

Chapitre 8 : Les sciences de la reconnaissance 99

Chapitre 9 : Le darwinisme neuronal 109

Chapitre 10 : Mémoire et concepts : un pont vers la conscience .. 131

Chapitre 11 : La conscience ou le présent remémoré 147

Chapitre 12 : Langage et conscience d'ordre supérieur..... 165

Chapitre 13 : L'attention et l'inconscient..... 181

Chapitre 14 : Les niveaux et les boucles : un résumé..... 193

QUATRIÈME PARTIE

Harmonies

Chapitre 15 : La philosophie et ses affirmations : un cimetière aux « ismes » 205

Chapitre 16 : La mémoire et l'âme : non au réductionnisme idiot 217

Chapitre 17 : Pensées, jugements, émotions : des produits supérieurs 227

Chapitre 18 : Maladies mentales : le moi réintégré 235

Chapitre 19 : Peut-on construire un objet conscient ?..... 249

Chapitre 20 : Symétrie et mémoire : aux sources de l'esprit 261

Épilogue 277

L'esprit sans la biologie : postface critique..... 279

Bibliographie..... 333

Sources des citations et des illustrations 351

Index thématique..... 357

Remerciements 363

Liste des illustrations..... 365

· · · SAGIM · CANALE · · ·

Achevé d'imprimer en février 2008
sur rotative Variquik
à Courtry (77181)

Imprimé en France

Dépôt légal : février 2008
N° d'impression : 10662
N° d'édition : 7381-2071-X

L'imprimerie Sagim-Canale est titulaire de la marque
Imprim'vert® 2007

Bibliothèque

Gerald M. Edelman
Biologie de la conscience

Comment pensons-nous ? Qu'est-ce qui fait de nous des êtres doués de conscience, capables de nous souvenir, de percevoir le monde alentour, d'éprouver des passions ?
Ce livre présente l'ensemble des mécanismes qui composent l'esprit humain et dresse le bilan de la révolution accomplie par les neurosciences : la biologie du cerveau et l'étude de son évolution sont en passe de nous fournir la clé de la conscience elle-même.

« Une ouverture sur l'une des voies scientifiques les plus passionnantes des années à venir. » *Le Nouvel Observateur.*

« Un livre des plus profonds parmi les ouvrages sur ce sujet capital. » *Études.*

« Un magistral exposé et une superbe réflexion sur l'homme. » *Paris Match.*



Prix Nobel de physiologie et de médecine, Gerald M. Edelman est directeur du Neurosciences Institute, à La Jolla, en Californie, président de la Neurosciences Research Foundation et chef du département de neurobiologie du Scripps Research Institute. Il est l'auteur de *Comment la matière devient conscience*, *Plus vaste que le ciel* et de *La Science du cerveau et la connaissance*.

9 782738 120717
713709.3
ISSN 1957-9411
ISBN 978-2-7381-2071-7
En couverture : *In double minds*, technique mixte sur papier de Eva Eun-Sil Han.
4e de couverture : © Emmanuel Robert.

19,90 €

www.odilejacob.fr