

Paru dans Andler, D., dir., *Introduction aux sciences cognitives*,
Paris : Gallimard, nouv. éd. 2004, pp. 699-714

CHAPITRE 20

Les sciences cognitives à l'aube de leur deuxième demi-siècle

Daniel Andler

Table du chapitre

I. Des champs nouveaux

A. LA COGNITION DE L'AGENT SOLITAIRE

- (a) *Minimalisme et théorie de l'optimalité*
- (b) *Les concepts et leur acquisition; contenu conceptuel vs non conceptuel*
- (c) *L'agentivité : corps propre et contrôle de l'action*
- (d) *Le soi, la première personne, la mémoire*
- (e) *La conscience : théories neurobiologiques, analyses philosophiques*
- (f) *Modularité générale versus constructivisme neuronal ; pour ou contre l'innéisme*

B. LA DIMENSION SOCIALE DE LA COGNITION

- (a) *Cognition sociale : les bases psychologiques de notre aptitude à vivre avec autrui*
- (b) *Cognition située : les facteurs extra-individuels dans la cognition*
- (c) *Vers une théorie des cultures en forme de psycho-anthropologie évolutionniste*
- (d) *Cognition animale et approches comparatives; la question du langage*
- (e) *Economie : les courants expérimentaux et cognitifs se rejoignent*
- (f) *Emotions et affects. La cognition chaude est de retour*

C. LES « TECHNOLOGIES CONVERGENTES » : « NBIC » ET LES SCIENCES COGNITIVES APPLIQUÉES

- (a) *Les quatre protagonistes et la question de la technologie*
- (b) *Les liens entre C et N, B, I*
- (c) *Le retour de l'IA prométhéenne*
- (d) *Science et technologie*
- (e) *Sciences cognitives appliquées et sciences de la cognition appliquée*
- (f) *L'interdisciplinarité*

II. Le nouveau visage des sciences cognitives

A. LES NEUROSCIENCES, VAISSEAU-AMIRAL?

- (a) *L'irrésistible ascension de la neuroimagerie*
- (b) *Critiques et objections*
- (c) *Au-delà de l'imagerie*
- (d) *Résultats et promesses : les leçons de l'IA*

B. OÙ EST PASSÉ LE MAINSTREAM?

- (a) *Les difficultés philosophiques du fonctionnalisme*
- (b) *La cause et le calcul*
- (c) *La multiplication des écoles, ou l'éclosion de mille fleurs*
- (d) *Formes et limites du naturalisme.*

BIBLIOGRAPHIE

I. Des champs nouveaux

Comme on l'a dit dans l'avant-propos, les candidats à la dignité de chapitre manquant sont nombreux. Dans cette première partie sont énumérés, et brièvement décrits, un choix de candidats peu contestables. Je reviendrai, dans la seconde partie, sur l'évolution des grandes disciplines et sur l'émergence de nouvelles méthodologies, qui motiveraient, par exemple, des chapitres métathéoriques sur "les nouvelles neurosciences cognitives", sur "les nouvelles méthodes d'imagerie fonctionnelle", sur "les nouvelles figures de l'intelligence artificielle", sur "le retour de la *Gestalt*" ou encore sur "le paradigme dynamique". Je m'en tiens pour le moment à des secteurs de recherche : nouvelles théories, nouveaux objets. L'échantillon que je propose n'est par définition, pas exhaustif, et je me garde d'affirmer que chacun s'accorderait à le trouver suffisamment représentatif. J'ai choisi douze questions qui n'auraient pas pu figurer, du moins sous leur forme théorique présente, dans le volume de 1992, et qui figureraient très probablement, sous des intitulés ou dans des découpages peut-être différents, dans un recueil conçu aujourd'hui. Ces douze questions se partagent, à égalité, entre des propriétés du système cognitif qui ne sont pas directement liées, en première analyse, à l'environnement social de l'individu, et des propriétés qui le sont. Cette répartition est en elle-même révélatrice d'un déplacement du centre de gravité du domaine, d'autant que même les questions de cognition "non relationnelle", pour le dire brièvement, apparaissent souvent comme indirectement liées à la cognition "relationnelle" : les sciences cognitives sont en passe de devenir une composante essentielle des sciences sociales. Il ne faut pas voir dans cette dichotomie une hypothèse implicite ; elle est commode, mais il ne s'agit pas ici de la défendre sur le plan théorique, d'autant qu'elle se heurte à l'une des idées-forces qui circule beaucoup actuellement, selon laquelle la cognition n'est jamais intégralement individuelle. Enfin, j'évoquerai rapidement les sciences cognitives appliquées, qui sont partie prenante des "technologies convergentes".

A ces douze questions viennent s'ajouter des sujets traités dans le corps du volume et qui ont connu un développement considérable. Je pense notamment à la sémantique et à la pragmatique, dont certains aspects sont présentés dans les chapitres 7 et 8 : ces domaines ont beaucoup évolué tant dans le sens de la formalisation que sous l'effet de l'expérimentation, et leur rôle va croissant (notamment à cause d'internet). D'autres thèmes seront évoqués dans la seconde partie du chapitre.

A. LA COGNITION DE L'AGENT SOLITAIRE

L'époque héroïque est marquée, comme le rappelle l'*Introduction* du présent volume ("Le cognitivisme sybillin", p. 40), par certaines hypothèses auxiliaires et certaines exclusives, qui vont au-delà de l'hypothèse cognitiviste centrale, et permettent à celle-ci de se déployer dans des programmes de recherche. Ce sont autant de manières de spécifier conjecturalement le cadre conceptuel classique, et la nouveauté des questions de notre première liste consiste à mettre en question, de façon plus ou moins radicale, certaines de ces spécifications.

(a) *Minimalisme et théorie de l'optimalité*. La grammaire générative, dont le projet est conçu dès le milieu des années 1950 par Noam Chomsky, a été durablement liée, d'une part, au cognitivisme

classique (dont la formule canonique constitue le titre de l'un des ouvrages du linguiste : *Règles et représentations*), et d'autre part à une conception particulière de l'objet d'une science linguistique, articulant la recherche d'une description des langues et l'explication de leur acquisition par tout enfant normal. L'entreprise a d'abord pris la forme d'une spécification de la grammaire universelle, des grammaires des langues naturelles et des transformations qui font passer d'une structure profonde à une structure de surface des phrases. Depuis une vingtaine d'années, Chomsky a cherché à simplifier la théorie, en se plaçant à un niveau d'abstraction plus élevé, d'où tout concept substantiel de règle, et partant de "structure profonde" et "structure de surface", sont exclus. L'unique contrainte réelle à laquelle est soumise la faculté de langage est celle qu'exercent ses interfaces avec, d'une part, le système intentionnel-conceptuel, et d'autre part le système sensori-moteur. Autrement dit, le langage a pour fonction de traduire un input constitué d'une pensée conceptuelle en un output constitué d'une suite de sons. Une théorie du langage au sens strict n'a pour parties que les interfaces, forme phonétique d'une part, forme logique de l'autre. Ainsi se dessine, depuis le début des années 1990, le programme minimaliste, avant toute hypothèse plus spécifique sur les théories linguistiques susceptibles de réaliser ce programme.

A peu près dans le même temps, Paul Smolensky engageait avec Alan Prince, l'un des linguistes qui formula, il y a une quinzaine d'années, des critiques théoriques sur le connexionnisme de l'école PDP, une discussion dont allait bientôt émerger une théorie qui partage avec l'hypothèse minimaliste le souci de généralité théorique, et la recherche d'une authentique synthèse entre la linguistique théorique et les sciences cognitives. La théorie de l'optimalité conserve, de la linguistique générative classique, non seulement la conception de son objectif central, qui est de décrire la grammaire universelle, mais aussi la notion de règle. Son originalité réside dans sa manière de concevoir cette dernière notion et de l'intégrer dans le cadre d'une théorie de l'harmonie généralisée (v. chap. 2, III), conçue comme modalité fondamentale du fonctionnement de l'appareil cognitif, susceptible tant d'une caractérisation psycholinguistique que d'une description neurobiologique. Les règles sont des contraintes universelles, mais elles sont conflictuelles et les différentes langues sont autant de solutions au problème de minimiser les violations; elles diffèrent par l'ordre de priorité qu'elles assignent aux contraintes. Comme pour le minimalisme, il ne s'agit que d'un cadre, et le vrai travail consiste à rendre compte, progressivement, des régularités linguistiques, d'une part, des processus d'acquisition, d'autre part.

Ces deux théories, venant d'horizons si éloignés, convergent à deux niveaux : celui de l'objet d'étude – elles portent sur le même objet –, et elles se réclament d'une certaine exigence théorique, d'un "esprit scientifique" qui n'est pas sans rappeler la biologie théorique telle que la concevait un René Thom.

(b) *Les concepts et leur acquisition; contenu conceptuel vs non conceptuel.* Les concepts sont évidemment la notion centrale de toute la théorie de la connaissance depuis les Anciens, ils se trouvent à l'origine de la philosophie du langage issue de Frege, ils sont à la charnière de celle-ci et de la philosophie des sciences; enfin, ils figurent en bonne place dans la philosophie de l'esprit née dans les années 1960, en relation avec l'innéisme néocartésien de Chomsky, avec l'hypothèse du langage de la pensée exposée par Fodor, avec les recherches de Gareth Evans ou Christopher Peacocke sur les variétés de la représentation et le non-conceptuel, avec la critique d'inspiration phénoménologique faite par Dreyfus d'un fondement conceptuel de la cognition, avec la question de

l'imagerie mentale telle qu'elle est posée au début des années 1980, etc. On ne peut donc surestimer leur rôle permanent en philosophie.

Pour leur part, les psychologues ont fait depuis longtemps des concepts un thème de recherche important. C'est à l'acquisition par étapes des concepts logiques, mathématiques, spatiaux, temporels,... (bref, l'appareil de la "raison pure") qu'est consacrée une partie essentielle de l'œuvre de Piaget. De nombreux travaux de la psychologie cognitive naissante et de l'intelligence artificielle tournent autour des mêmes questions, cette fois dans la perspective d'un système cognitif adulte possédant, ou acquérant, des concepts spécialisés. Dès le début des années 1970, Eleanor Rosch proposait une révision radicale de la théorie plus ou moins implicitement admise par les psychologues, selon laquelle un concept se définit par un ensemble de conditions nécessaires et suffisantes, et sa maîtrise par la connaissance de ces conditions.

Pourtant, pour des raisons qui seraient intéressantes à élucider, tout se passe comme si philosophes et psychologues s'étaient éveillés, au cours des dix ou douze dernières années, d'une sorte de sommeil dogmatique, et s'étaient rendu compte de la nécessité de reprendre à nouveaux frais la question des concepts. Cette redécouverte de questions qu'en un sens on n'avait jamais perdues de vue, mais dont on n'avait pas perçu certaines dimensions, et certains enjeux, s'est manifestée sur plusieurs terrains. Les philosophes ont voulu mieux comprendre comment articuler une philosophie de l'esprit compatible avec les sciences cognitives contemporaines, d'un côté, avec les conceptions philosophiques courantes des concepts, de l'autre, et ils ont pris conscience de la difficulté du problème; le titre d'un livre récent de Fodor l'exprime sans détour : *Concepts. Where cognitive science went wrong* ("Les concepts, ou comment les sciences cognitives se sont fourvoyées"). Les psychologues du développement, reprenant le programme piagétien avec des outils et des présupposés théoriques très différents, ont commencé à explorer le "bagage conceptuel" des nourrissons et des jeunes enfants, et ont proposé des hypothèses rivales sur le changement conceptuel, c'est-à-dire sur le passage d'une forme de conceptualisation (implicite) d'un domaine (celui des nombres, par exemple, ou des espèces vivantes, ou de la mécanique des objets courants). Les psychologues expérimentaux (travaillant sur la cognition de l'adulte) ont considérablement approfondi leurs théories "opératoires" des concepts, et abouti à trois conceptions rivales (théories, prototypes et exemplaires) qu'ils tentent de départager, ou bien alors de combiner, dans une débauche d'expériences parcellaires. Les neuropsychologues (travaillant sur l'adulte ou l'enfant cérébrolésé) proposent des découplages inattendus entre capacités conceptuelles de différents ordres. Les philosophes et les neurobiologistes s'intéressant à la perception et à la motricité s'interrogent sur la notion de contenu non conceptuel des représentations hypothétiquement impliquées dans ces processus, et sur le rapport entre ce contenu et le contenu des représentations conceptuelles. Les psychologues sociaux se demandent de quels types de concepts sont constitués les « représentations » (de soi, d'autrui, du groupe, des autres groupes...) qui sont leur objet central. Enfin, ces différents développements, par leur importance même et leur logique propre, posent la question de leurs rapports mutuels : parle-t-on réellement de la même chose dans ces différents programmes de recherche? Certains en doutent, et l'on peut se demander ce que va devenir au cours des années qui viennent l'idée même de concept dans les sciences cognitives et la philosophie de la connaissance.

Et pourtant, si les concepts sont remisés, que devient la *pensée conceptuelle*? Une forme de pensée dont on ne peut sûrement pas se passer : c'est, en gros, celle qui rend possible la réflexivité, dont il est difficile de contester le rôle central dans la rationalité. Une famille d'états mentaux

caractéristiques de la pensée conceptuelle est constituée par les attitudes propositionnelles : Dès que je crois qu'*il est dix heures*, je peux procéder à toute une série de transformations logiques et dialectiques qui étendent mon horizon épistémique bien au-delà de la reconnaissance brute du fait. Je peux penser qu'il n'est pas neuf heures. Je peux essayer de convaincre mon voisin qu'il est dix heures. Je peux prévoir que dans une heure il sera onze heures (si je dispose également d'autres concepts, ou d'autres pensées conceptuelles). Si nous admettons la notion de pensée conceptuelle, nous devons admettre les concepts, mais nous pouvons le faire de façon définitionnelle, et dire que les concepts sont des constituants des pensées conceptuelles, au lieu de chercher à caractériser les concepts indépendamment, pour constater ensuite que les entités ainsi définies (peut-être comme des représentations mentales d'un certain genre, ou des dispositions à former de telles représentations) se trouvent entrer dans la composition des pensées conceptuelles.

Mais toute pensée n'est-elle pas conceptuelle ? Il y a au moins trois raisons de soupçonner que non. La première est que la perception et l'action mettent en jeu des pensées dont la caractérisation fait appel à des concepts que le sujet ne possède pas : un enfant, ou toute personne ignorante du tourisme parisien, peut percevoir la tour Eiffel sans avoir la croyance qu'il a sous les yeux la tour Eiffel ; un skieur peut imprimer à son corps un mouvement lui permettant de négocier une couche de neige fraîche sans pouvoir exprimer en mots ce qu'il fait. La seconde raison est que nous sommes le siège de processus dont les constituants élémentaires échappent à toute caractérisation conceptuelle (par exemple, lorsque nous élaborons une représentation tridimensionnelle d'un objet placé devant nous, en formant, si l'on en croit certaines théories, une succession de représentations intermédiaires indescriptibles dans le langage commun). La troisième raison est que les êtres vivants dépourvus de langage (animaux, bébés) semblent privés de pensée conceptuelle (ou du moins fortement limités sur ce plan), mais non privés de toute pensée. Ce que ces différentes situations ont en commun est qu'elles impliquent un agent dont les capacités représentationnelles excèdent ses ressources conceptuelles. On voit ainsi qu'il y a là des questions (une pensée non-conceptuelle est-elle possible ? est-elle possible en l'absence de *toute* pensée conceptuelle ? etc.) qui ne concernent pas seulement la philosophie de la connaissance, mais les sciences cognitives dans leur ensemble, ainsi que la philosophie morale.

(c) *L'agentivité : corps propre et contrôle de l'action.* Philosophes, physiologistes, neuropsychologues et neurobiologistes se sont penchés ensemble sur des phénomènes restés longtemps la province de la psychiatrie, de la phénoménologie et de leur entre-deux, et dont la particularité est de concerner le corps vivant, dans sa matérialité, un corps irréductible à un pantin articulé qui ne ferait que recevoir les ordres d'un centre de contrôle distant, informé par des capteurs sensoriels sur l'état de l'environnement.

On connaît depuis longtemps le phénomène du membre-fantôme (membre ou segment amputé qui se "signale" à la conscience par des sensations proprioceptives ou douloureuses), on sait aussi que la schizophrénie peut entraîner des phénomènes de brouillage de la perception du mouvement volontaire ("on a bougé mon bras") ou de la possession d'un membre ("ce bras n'est pas le mien"). Les notions de proprioception, de kinesthésie ou perception du mouvement, de mouvement propre et de corps propre sont d'autre part au cœur de tout un courant de la phénoménologie associé à Maurice Merleau-Ponty, mais qu'un Mach et des physiologistes tels que le Russe N. Bernstein ont partiellement anticipées ou développées pour leur propre compte. Ces idées sont souvent liées à la théorie motrice de la perception, selon laquelle la perception n'est

logiquement et empiriquement possible que pour un système doté de capacités motrices (“la perception par l'action” en est le slogan), théorie ancienne qui connaît un regain de faveur spectaculaire dû aux progrès des neurosciences (travaux de Milner et Goodale, notamment).

La nouveauté réside ici dans la conjonction de plusieurs mouvements d'idées, hypothèses et découvertes.

D'une part, la physiologie du mouvement, notamment en France dans les œuvres d'Alain Berthoz et de Marc Jeannerod, a quitté les marges des sciences cognitives pour en constituer, par le rapprochement avec la philosophie et la psychologie expérimentale, un des noyaux les plus vivants. L'agentivité est désormais une question centrale, liée à d'autres questions d'actualité, telles que la perception et le contenu non conceptuel, le soi, la conscience, la “psychologie naïve”, tous sujets dont on dira quelques mots plus loin. Les philosophes ont mesuré la difficulté d'une théorie de l'action motrice, et plus généralement l'insuffisance des schémas disponibles pour comprendre une action quelle qu'elle soit. Les physiologistes, à la suite des célèbres expériences de Benjamin Libet, ont mis au jour l'insuffisance de la conception traditionnelle, qui plaçait à l'origine d'un mouvement volontaire une décision consciente, suivie d'exécution.

D'autre part, les théories actuelles du contrôle du mouvement font appel à l'idée de simulation, voire à celle d'émulation. Les “ordres” donnés aux organes effecteurs seraient également transmis à un “simulateur” calculant (ou évaluant) les conséquences de ces ordres, lesquelles seraient comparées aux changements constatés, d'où s'ensuivrait un ajustement, et l'itération en temps réel de ces processus parallèles permettrait de rendre compte de l'extrême dextérité avec laquelle sont exécutés des mouvements fins et complexes. Telle est la forme générale d'une théorie de la simulation. Selon les défenseurs de l'émulation, un simple simulateur abstrait ne suffit pas : les ordres passés par le centre sont “entrés” dans un modèle interne du système effecteur, et les résultats sont “lus” (constatés) pour être ensuite comparés à l'état mesuré du système effecteur.

Enfin, la robotique, à la suite notamment des travaux de Rodney Brooks, a pris, sous l'étiquette de “behavior-based robotics” (BBR) un tournant “antireprésentationnel”, “incarné” (embodied), situé, évolutionniste et autonomiste qui entre en résonance avec les nouvelles orientations de la physiologie du mouvement biologique.

Quelles que soient les différences entre ces hypothèses et la diversité de leur provenance (philosophie, physiologie, neurosciences, modélisation...), s'en dégage une vision générale du mouvement qui lui confère une complexité et une autonomie par rapport au centre “cogitateur” de nature à modifier très profondément les conceptions traditionnelles des rôles respectifs de la perception, de la réflexion et de l'action.

(d) *Le soi, la première personne, la mémoire.* Proche du thème précédent, celui du soi et de ses attributs principaux s'en distingue par le fait qu'il le déborde : le soi n'est pas seulement un agent, l'auteur (problématique) de ses gestes et le possesseur de son corps. Il est aussi, au minimum, une mémoire d'un genre très particulier, une unité et une permanence perçues fondant la personne, le vecteur de la responsabilité, et enfin un point de vue privilégié, la fameuse “perspective à la première personne”. C'est sans doute le soupçon d'une irréductibilité non seulement ontologique, sur laquelle le philosophe Thomas Nagel avait attiré l'attention dès son célèbre article de 1974, mais également logique, sémantique et génétique, de cette “égoïté” fondamentale, qui a déclenché une réflexion nouvelle dans les sciences cognitives, lesquelles se sont ainsi rapprochées notamment de la psychanalyse, sans pour cela avoir encore su ou voulu réaliser avec elle une authentique jonction.

Linguistes, philosophes du langage et logiciens s'interrogent sur "l'indexicalité essentielle" (selon l'expression de John Perry), qui semble interdire le remplacement de toute une catégorie d'expressions et de représentations dans lesquelles figure une référence implicite ou explicite au sujet de l'énonciation ou au "possesseur" de la représentation par des expressions ou représentations synonymes dans lesquelles cette référence est absente.

Les philosophes reprennent à nouveaux frais la question de la mémoire et de sa "possession" par celui qui se souvient. A la suite du philosophe Derek Parfit, ils comparent la mémoire à la "quasi-mémoire", dont le contenu est identique à celui d'une mémoire ordinaire, mais dont l'attache au soi est absente. Les psychologues étudient la mémoire autobiographique, le phénomène des personnalités multiples et la dépersonnalisation du schizophrène ("j'ai pensé que ceci ou cela, mais ce n'est pas moi qui pense ceci ou cela; ce n'est pas *ma* pensée"). Les spécialistes du développement rejoignent les philosophes dans une mise en cause d'une différence d'essence entre l'accès que le sujet a à ses propres états mentaux et l'accès aux états mentaux d'autrui, en se fondant sur l'étude de l'acquisition des concepts correspondants chez le jeune enfant (nous y reviendrons ci-dessous, en 1.2.a). Les bases de la connaissance de soi semblent moins fermes qu'on ne le pensait.

De même, la notion de personnalité, ensemble solidaire de traits de caractère durables, fait l'objet de questionnements plus ou moins radicaux. L'intuition qui nous guide dans l'interprétation des comportements d'autrui ou de nous-même est que la personnalité de l'agent est le déterminant essentiel. Or de nombreuses expériences semblent montrer que souvent la situation l'emporte sur tout autre déterminant : nous commettrions, selon cette analyse, l'"erreur fondamentale d'attribution" désormais bien repérée dans la psychologie sociale. Certains, comme le jeune philosophe John Doris, vont plus loin : ils se demandent s'il ne faut pas tout simplement éliminer de l'ontologie scientifique toute notion de personnalité.

Remarquons qu'avec cette question nous pénétrons dans le domaine de la cognition sociale, franchissant la frontière un peu arbitraire entre les deux moitiés de notre liste, mais aussi la frontière institutionnelle bien réelle qui a durablement isolé les sciences cognitives de la psychologie sociale. La notion de personnalité est indissociable de celle de la perception ou de l'attribution de personnalité, et elle voisine avec d'autres thèmes caractéristiques de la psychologie sociale, on y reviendra plus loin (1.2.b).

(e) *La conscience : théories neurobiologiques, analyses philosophiques.* Intimement lié aux deux thèmes précédents, celui de la conscience est certainement le plus caractéristique du nouveau visage des sciences cognitives. Après avoir été, pour des raisons diverses, exclu du champ, et avoir connu une traversée du désert longue d'environ un demi-siècle, la conscience est revenue, depuis une dizaine d'années, par la grande porte : philosophes et neurobiologistes se disputent la faveur d'en parler (la *MITECS* résout le problème en publiant deux articles distincts); c'est le sujet-vedette, le sujet des vedettes, une spécialité à soi seul, avec revues, congrès spécialisés, haut et bas clergé...

Les philosophes s'efforcent de clarifier la notion, et d'en distinguer différentes strates. A titre d'exemple, et sans vouloir présenter un tableau même succinct de la situation conceptuelle, le philosophe Ned Block a proposé de distinguer deux formes de conscience (précisons néanmoins que l'on écarte, dans cette discussion, l'idée de conscience associée à l'état de veille, et l'idée de conscience morale). La conscience d'accès (A-conscience) entretient avec un contenu associé un certain type de rapport : la perception A-consciente de la pomme permet au sujet de procéder à toute une série d'inférences quant à la présence, devant lui, de la pomme; de même le souvenir A-conscient

de la pomme posée sur l'assiette, la pensée A-consciente qu'une pomme s'achète au marché, etc. La conscience phénoménale (P-conscience) entretient avec un contenu associé un autre type de rapport, un rapport qualitatif, non inférentiel : la perception P-consciente de la pomme consiste en une sensation distinctive, un "ce-que-ça-fait" de mordre dedans, ou simplement de la voir sur l'assiette. Les deux types de conscience ne s'appellent pas nécessairement l'une l'autre : on peut peut-être éprouver un "ce-que-ça-fait" caractéristique de mordre dans une pomme sans pouvoir exploiter cognitivement le fait que l'on est en train de mordre dans une pomme; inversement, on peut sans doute savoir qu'on mord dans une pomme sans éprouver le "ce-que-ça-fait" correspondant.

Les philosophes s'interrogent néanmoins sur le rapport entre les deux formes de conscience, sur leur réalité respective. Certains jugent que la P-conscience n'est pas moins représentationnelle que l'autre, mais qu'elle implique des représentations particulières. D'autres pensent au contraire que la P-conscience est une notion irréductible, qui pose à l'ontologie et à la science un problème absolument unique, et peut-être insoluble.

Philosophes et psychologues se demandent aussi à quoi sert la conscience, sur le plan cognitif. A-t-elle une fonction, et en particulier peut-on s'expliquer son apparition au cours de l'évolution ?

Les neurobiologistes sont dans l'ensemble bien moins précautionneux. D'abord, certains estiment avoir identifié un mécanisme neuronal sous-jacent à la conscience : il s'agirait d'une synchronisation des activités de certains neurones sur une certaine fréquence spécifique, de l'ordre de 40 Hz. D'autres élaborent des théories plus complexes, qui visent également à isoler un ensemble de cellules et de processus constituant le "corrélât neuronal de la conscience". Ensuite, on trouve des explications très générales, proposées par des biologistes de renom, sur la base de leur compréhension de l'architecture cérébrale. En réalité, les idées précises ne concernent qu'un aspect de la conscience, la vision consciente, et ne l'"expliquent" guère, comme le reconnaissent avec simplicité Christof Koch et Francis Crick. Quant aux schémas globaux, ils relèvent davantage de spéculations et de "suggestions" que de théories au sens scientifique ou philosophique du terme; cela ne leur ôte évidemment pas toute valeur, à condition qu'elles soient prises pour ce qu'elles sont.

La situation est très particulière, du fait que l'idée même d'expliquer la conscience semble difficile à saisir. Pour certains, tels Dennett, ce que nous devons expliquer, ce n'est pas la conscience, mais l'illusion où nous sommes qu'elle existe. D'autres parlent du "fossé explicatif", pressenti par Nagel, et que Joseph Levine a baptisé et cerné : quelle que soit la "théorie" que pourront nous proposer demain la neurobiologie (Edelman, Damasio, Crick et Koch, Dehaene,...) ou la physique quantique (Penrose...), nous ne comprenons pas aujourd'hui en vertu de quoi elle pourrait être considérée comme une théorie explicative de la conscience. Le vrai mystère, ce que David Chalmers appelle le "problème difficile", est celui de la place de la P-conscience dans le monde naturel. Certains imaginent d'échapper à l'aporie en affirmant une thèse de pure identité : la conscience *n'est rien d'autre qu'une propriété neurobiologique*; cette radicalisation de la position matérialiste contemporaine (tout phénomène psychique est *aussi* un phénomène cérébral, au sens minimal d'être en quelque sorte dépendant d'une trace dans le tissu neural) est la symétrique de ce nouveau dualisme dont Chalmers est le champion (aucune explication neurobiologique ne *peut* nous éclairer sur le phénomène de la conscience).

Rares cependant sont ceux qui contestent que la problématique évolue sous l'effet des progrès de la neurobiologie, et en particulier grâce à l'apport de la neuropsychologie. Un phénomène tel que celui de la "vision aveugle" (*blindsight*), décrit dès 1905 par le neurologue suisse L. Bard,

redécouvert et théorisé dans le contexte actuel par Lawrence Weiskrantz, ébranle durement nos intuitions : comment peut-on avoir une perception visuelle, si appauvrie soit-elle, en ayant la conviction absolue de n'en avoir aucune?

(f) *Modularité générale vs constructivisme neuronal ; pour ou contre l'innéisme.* Les trois derniers thèmes qu'on vient de passer en revue étendent le domaine des sciences cognitives au-delà des frontières qu'elles s'étaient provisoirement données dans la première phase de leur développement. Nous regagnons maintenant le cœur historique du domaine : l'architecture de l'esprit connaissant, les fondements naturels de cette architecture.

Jerry Fodor, dans son ouvrage historique de 1982 *la Modularité de l'esprit*, donnait une interprétation contemporaine d'une distinction traditionnelle entre fonctions inférieures (partagées avec les animaux — l'expression politiquement correcte dans le milieu et à mon sens ridicule étant "animaux non humains") et fonctions supérieures : les premières, rebaptisées *systèmes d'entrée* ("input systems"), posséderaient des caractéristiques computationnelles et informationnelles les prêtant à la théorisation par les moyens actuels des sciences cognitives; les secondes, regroupées sous l'étiquette de "processus centraux", privées de ces caractéristiques, présenteraient aujourd'hui, et sans doute à jamais, des obstacles insurmontables par cette science. (Le langage serait quant à lui très cartésienement partagé entre les deux familles —une partie automatique, rapide, irrépessible, quasi-autonome, et une partie interprétative aux connexions multiples à d'autres facultés et sources d'information.) D'un côté, une nature suffisamment organisée pour se laisser prendre aux filets de la science, visiblement régie par l'évolution, inscrite dans le patrimoine génétique; de l'autre, un magma insaisissable, quoique sans doute inclus dans l'ordre naturel, mais trop désordonné pour faire l'objet d'une science.

Si les années 1980 ont été celles de la modularité à la Fodor, qu'on pourrait qualifier, a posteriori, de « restreinte », les années 1990 voient s'affirmer la thèse rivale de la modularité "générale" (on dit plutôt "massive"). Les processus centraux de Fodor forment eux aussi, selon cette thèse, un ensemble structuré de capacités distinctes, dont chacune est spécialisée dans le traitement d'un certain domaine de l'environnement naturel. La "fixation de la croyance" n'est pas l'affaire d'une capacité généraliste, mais de capacités spécialisées, "*domain-specific*": les objets de taille et forme ordinaire, leur manipulation et leurs trajectoires forment le domaine de la "physique naïve"; nos congénères en tant qu'agents, munis chacun de croyances et de désirs qui leur sont propres et expliquent leur comportement, sont le domaine de la "psychologie naïve"; les espèces et variétés animales et végétales, celui de la "taxinomie naïve"; les nombres, celui d'une "arithmétique naïve"; les groupes sociaux et leurs interrelations sont, peut-être, le domaine d'une "sociologie naïve", etc. Ces capacités se développent à partir de noyaux de connaissance implicite que les psychologues du développement commencent à circonscrire et appellent "*core knowledge*". Ces modules "supérieurs", comme les modules fodorien, sont des produits de l'évolution et possèdent une base innée, responsable de structures cérébrales repérables. Dans l'environnement actuel, qui diffère à de nombreux égards de celui dans lequel, à l'âge du pléistocène, l'évolution a mis la dernière main à *Homo sapiens sapiens*, il arrive que ces modules traitent des informations qu'ils n'ont pas été "conçus" pour traiter de manière adaptative, et donnent lieu à des dysadaptations ou à des exaptations.

Ainsi brossé à très grands traits, ce programme de recherche de la "*domain-specificity*", (locution que j'ai proposé de traduire par "domanialité") pose à nouveaux frais la question de l'innéisme, et rencontre ainsi les problématiques actuelles de la théorie de l'évolution. Une nouvelle

interface s'est ouverte, qui constitue l'un des foyers les plus actifs des sciences cognitives, rapprochant psychologie cognitive et développementale, psychologie évolutionniste, théorie de l'évolution, neurosciences (du développement du système nerveux à la neuropsychologie), génétique, paléontologie, philosophie des sciences, et, nous allons le voir, sciences sociales.

A ce programme tend à s'opposer le "constructivisme neuronal" qui s'appuie sur la considération de certaines propriétés du cerveau pour s'opposer aux idées associées d'innéisme et de domanialité. La plasticité du cerveau, son développement anté- et postnatal, sa capacité de créer des systèmes de représentations flexibles adaptés aux tâches rencontrées, son autonomie, tout cela, aux yeux de ces nouveaux constructivistes, affaiblit considérablement l'argument-clé des innéistes, la "pauvreté du stimulus". Contrairement à ce que ceux-ci affirment, du fait que l'information reçue par l'enfant est trop restreinte pour permettre à un système computationnel d'apprentissage général, non biaisé, d'identifier la structure du domaine à acquérir, on ne peut, selon les constructivistes, déduire qu'une partie essentielle de l'information est présente au départ (thèse de l'innéisme), car une telle inférence repose sur une notion de calculateur fixe incompatible avec les enseignements des neurosciences. (Si ce qui précède semble trop abstrait, on pourra relire l'article de Daniel Osherson dans le présent volume .)

Avec cette controverse, nous touchons à la question des rivalités de disciplines et d'options théoriques, que j'évoquerai plus loin. Il est à peu près clair que les arguments des néoconstructivistes "neuronaux" sont soit trop fragiles pour réellement ébranler les tenants de l'innéisme en matière de langage, soit trop généraux pour ébranler les défenseurs de la psychologie évolutionniste, mais qu'ils ne sont pourtant pas sans portée. Chaque camp, installé au cœur d'une discipline, s'appuie sélectivement sur certaines données des autres disciplines pour se prévaloir d'une conception unifiée, du moins à terme. Il est beaucoup trop tôt pour espérer trancher. En attendant, il y a du pain sur la planche des deux côtés.

B. LA DIMENSION SOCIALE DE LA COGNITION

La locution "cognition sociale" tendant à se répandre, pourquoi ne pas l'adopter comme titre de cette deuxième rubrique? Simplement parce qu'elle est ambiguë. Au sens étroit, elle désigne les capacités cognitives fondamentales des individus qui leur permettent de participer aux rapports sociaux. Au sens large, elle inclut également des processus collectifs méritant, à divers titres, d'être considérés comme cognitifs. Nous allons passer en revue, successivement, des exemples des deux genres – en (a) et (b) – avant d'aborder des thèmes qui relèvent à la fois de l'un et de l'autre.

(a) *Cognition sociale : les bases psychologiques de notre aptitude à vivre avec autrui.* Commençons par la fin, qui est le début dans l'ordre de la découverte : la désormais célèbre "théorie de l'esprit". Le primatologue David Premack pose dès 1978 (dans un article aujourd'hui classique avec G. Woodruff) la question de savoir si les chimpanzés possèdent ou non la capacité de prédire le comportement d'un congénère (ou d'un homme) à partir d'"hypothèses" sur ses croyances et ses désirs : si l'autre veut X (manger la banane) et croit Y (que la banane est cachée derrière l'arbre), alors il essaiera (sauf circonstances adverses) d'aller voir derrière l'arbre. Cette capacité, qui repose sur l'attribution à l'autre de croyances et de désirs (et le cas échéant d'autres attitudes telles que l'espoir, la crainte, la préférence, l'attente,...) lui appartenant en propre dans la situation considérée et

principiellement distincte des miennes (je ne veux pas nécessairement manger la banane, et je ne crois pas nécessairement qu'elle soit cachée derrière l'arbre), et sur l'application d'un syllogisme pratique à ces prémisses, est ce que désigne la "*theory of mind*" ou ToM, théorie de l'esprit (TDE), une expression à prendre donc en un sens extrêmement particulier, ou encore "psychologie naïve", à entendre également en un sens technique. (TDE est le terme préféré des psychologues, psychologie naïve ou *folk psychology* celui des philosophes.)

La nature de cette capacité, ses constituants fondamentaux, son mode d'existence tant chez les adultes que chez les jeunes enfants normaux (qui l'acquerraient, semble-t-il, entre 3 ans et 5 ans) ou atteints d'autisme (qui ne l'acquerraient pas), constituent désormais un champ de recherche à part entière. On se demande si cette capacité est de l'ordre d'une "théorie" (c'est la "théorie de la théorie") ou repose sur la simulation (selon la "théorie de la simulation", pour prédire ce que l'autre va faire, je me mets à sa place, m'attribuant provisoirement ses croyances et ses désirs, et je "lis" en moi-même ce que je suis porté à faire). La seconde hypothèse soulève la question, cruciale en elle-même, de savoir si j'acquiers la connaissance de mes propres représentations d'une façon fondamentalement différente ou au contraire semblable à celle dont j'acquiers la connaissance des représentations de l'autre.

Relèvent également du domaine de la psychologie naïve deux séries de questions sur l'amont et sur l'aval de la TDE.

En amont, on recherche les capacités élémentaires sur lesquelles repose la TDE elle-même : reconnaissance et discrimination des visages; reconnaissance de l'expression faciale des émotions élémentaires; repérage de la direction du regard de l'autre; attention conjointe (tu m'invites, par la direction de ton regard, à regarder ce que tu regardes; ainsi, toi et moi regardons la même chose et nous savons, toi comme moi, que nous regardons la même chose); sensibilité au mouvement biologique, c'est-à-dire auto-commandé, par opposition au mouvement mécanique réactif; perception de l'agentivité, chez moi et chez l'autre ("je lève mon bras", par opposition à "mon bras est soulevé par l'autre"; "il lève son bras", par opposition à "son bras est soulevé par l'action d'un tiers"); imitation. Une découverte qui a beaucoup frappé les esprits est celle des *neurones-miroirs* : Giacomo Rizzolatti et son équipe à Parme ont identifié en 1992 dans le cortex prémoteur du macaque des neurones qui sont sélectivement activés, indifféremment, lorsque l'animal tend la main pour saisir des cacahuètes qu'on lui offre ou lorsqu'il est témoin de la même scène, avec un expérimentateur jouant le rôle qu'il jouait lui-même. Ces neurones constitueraient le lien primitif entre mon action et l'action de l'autre, et seraient même, selon Vittorio Gallese, le plus philosophiquement ambitieux de l'équipe, rien de moins que la base matérielle de l'intersubjectivité.

En aval, il s'agit de comprendre si (et en quel sens précis) la TDE est le fondement du rapport à autrui, voire de la notion de personne ou de la notion de soi, c'est-à-dire rien de moins que le fondement individuel de la socialité. Quel est le rôle des émotions? Qu'il s'agisse de l'émotion sociale la plus simple, l'empathie (dont il faut soigneusement distinguer la variété imitative et contagieuse, empathie dite "instructive" - de te voir pleurer, ou rire, me pousse à pleurer, ou à rire -, de la variété « intentionnelle », consistant en une réaction adaptée - de te voir pleurer m'incite à te consoler), ou des émotions complexes qui mettent en jeu, et régulent, les rapports sociaux. Comment s'articule la moralité dans l'ordre naturel de ces structures psychologiques? La sociopathie, la psychopathie, sont-elles dues à une altération de la TDE ?

(b) *Cognition située : les facteurs extra-individuels dans la cognition.* La notion même de cognition est liée à celle d'un contexte ou d'une situation hautement variable : une créature dont l'existence se déroule dans un cadre fixe ou dont la variabilité est faible n'a pas besoin de pensée ni de système nerveux central (Dennett cite l'exemple de ce mollusque qui dérive jusqu'à trouver un rocher convenable auquel s'amarrer et dévore alors ce qui lui avait jusque-là tenu de cerveau). Les sciences cognitives n'ont en ce sens jamais fait autre chose que d'étudier des systèmes capables de remplir leur rôle en tenant compte du contexte. Pourtant, la notion de contexte a pris ces dernières années une place centrale, d'une part en raison des problèmes conceptuels qu'elle pose, d'autre part du fait qu'elle prend dans différents programmes de recherche des formes concrètes qui donnent lieu à des études empiriques ou à des modélisations.

Sans qu'on puisse aborder ici la question de savoir si c'est le fait d'une simple coïncidence, il faut constater que les problèmes de contexte les plus aigus et les plus étudiés font intervenir la dimension sociale.

En premier lieu, la cognition semble souvent, en un certain sens, distribuée entre différents agents. C'est une banalité de dire qu'une équipe de rugby "sait" davantage que ne "savent" ses différents membres. C'est une évidence que chacun, dans son quotidien comme dans sa pratique professionnelle, se repose sur de vastes quantités de connaissances, de compétences et d'actions qui sont détenues ou déclenchées par d'autres que lui. Qu'en particulier la connaissance scientifique, même limitée à une spécialité, n'est condensée dans l'esprit d'aucun savant, mais se distribue sur la communauté des experts. Ce truisme qui ne date pas d'hier, mais dont chacun constate la vérité toujours plus éclatante, conduit à deux niveaux d'interrogation. D'une part, on peut chercher à rendre compte des relations entre connaissances individuelles et connaissances détenues ou déployées par le groupe social. La pragmatique, et tout particulièrement la théorie de la pertinence de Sperber et Wilson, reposent sur l'idée que la compréhension d'une phrase, dans le contexte de la communication ordinaire, fait appel à des connaissances complexes impliquant solidairement les deux interlocuteurs. On élabore des systèmes de logique épistémique pour formaliser la dynamique des connaissances dans une collectivité d'agents. Edwin Hutchins a étudié empiriquement des situations hautement complexes telles que la navigation à bord d'un porte-avions ou le pilotage d'un avion gros porteur. L'informatique étudie le traitement distribué de l'information par un ensemble parfois considérable de processeurs semi-indépendants échangeant périodiquement des informations, résultats ou comptes rendus d'activité.

D'autre part, on peut poser la question ontologique : existe-t-il un niveau autonome de connaissance collective ou conjointe, quel est le statut des entités peuplant ce niveau, quel est leur rapport aux entités fondamentales de la cognition individuelle? Cette question est traitée par Dan Sperber dans son essai (chap. 15 du présent volume). Elle est intimement liée à ce qu'on appelle, dans la philosophie des sciences sociales, l'individualisme méthodologique. Rendre compatible cette doctrine et les phénomènes d'action conjointe sans perdre le bénéfice de l'approche naturaliste propre aux sciences cognitives est un problème difficile, qui exige d'affaiblir l'individualisme; à des degrés divers, D. Sperber, A. Nelson, J. Searle, M. Bratman, P. Livet, M. Gilbert suggèrent différentes manières de le faire.

Deuxième facteur de "distribution" : les artefacts, savoir-faire et institutions qui contribuent à l'exécution des tâches cognitives, et constituent ainsi non pas simplement des outils, mais davantage des "prothèses" cognitives. L'écriture est naturellement le premier exemple historique, et sans doute le plus central pour les sciences cognitives. L'aptitude à lire et à écrire est, d'une part, absolument

essentielle dans les processus cognitifs d'une majorité des humains, quoique absente chez beaucoup, et atteinte ou défectueuse chez certains (lésions neurologiques mais aussi dyslexies). D'autre part, elle est à la fois naturelle et non naturelle puisqu'elle n'est ni le résultat direct de la sélection naturelle, ni celui d'un processus individuel spontané, ni même enfin celui d'un processus social régulier et constant (l'invention d'un système d'écriture a été, comme le rappelle Jared Diamond, un événement extrêmement rare dans l'histoire de l'humanité). Dans le même ordre de phénomènes s'inscrivent les innombrables systèmes de repérage, d'inscription, de communication, les systèmes de documentation (à commencer par les bibliothèques, les index, les encyclopédies, aujourd'hui amplifiés et partiellement supplantés par les systèmes informatiques et le web) et de traitement de l'information (depuis les premières machines à calculer jusqu'aux ordinateurs individuels et en réseau). Mais il existe des supports externes moins visibles et néanmoins essentiels et omniprésents, qui consistent en aménagements de l'environnement matériel destinés à orienter le flux des opérations cognitives, depuis la disposition des manettes, interrupteurs et compteurs dans un poste de pilotage de machine-outil ou de véhicule jusqu'à celle des ustensiles et des ingrédients dans la confection d'un plat. Les sciences cognitives rencontrent ici la paléontologie, l'ethnologie et l'ethnométhodologie. En remontant dans l'histoire de l'humanité, la "culture matérielle" aurait, selon Steven Mithen et d'autres, permis au cerveau de s'étendre au-delà de ses frontières organiques, et de donner ainsi naissance à l'esprit, qui d'emblée aurait été de la sorte doté d'arc-boutants externes. Pour ce qui est de l'avenir, les sciences cognitives jouent un rôle capital, quoique encore insuffisamment reconnu, dans le vaste domaine des interactions ou interfaces homme/machine (HCI pour *human-computer interface*, terminologie qui tend à en masquer l'étendue) et de l'"utilisabilité" (on dit dans le jargon "usage", terme trompeur); on parle aussi de *cognitive engineering*, qui englobe l'environnement et le contexte de fins poursuivies dans lequel se situe l'interaction entre l'utilisateur et l'objet.

Enfin, comme on l'a dit plus haut, après des décennies de bouderie réciproque, sciences cognitives et psychologie sociale commencent à se rendre à l'évidence qu'elles ne peuvent plus continuer à s'ignorer, dans la mesure où, de part et d'autre, des phénomènes nouvellement abordés, ou repris à nouveaux frais, font appel aux compétences des deux traditions. Certes, le tournant cognitif de la psychologie sociale – qui historiquement prit précisément, dans les années 1960–70, le nom de "cognition sociale" – reste timide et relève d'un cognitivisme faible, consistant à introduire de-ci de-là quelques considérations sur les mécanismes cognitifs sous-jacents aux opérations de classification, appartenance, identification, transmission,... sociales, qui sont l'objet principal de cette discipline. Certes, inversement, les sciences cognitives ne font que commencer à aborder ces questions. Il n'empêche qu'on croise désormais, dans certaines recherches et dans certains ouvrages, des spécialistes des deux bords. Ainsi, l'investigation d'une "sociologie naïve" reprend la problématique de la psychologie sociale en se démarquant du courant dit de la "cognition sociale" (sans rapport direct avec les recherches évoquées au point (a) ci-dessus) par l'idée que les compétences sociales sont spécifiques ou domaniales : l'appréhension des groupes sociaux serait une capacité spécifique, largement isolée d'autres aptitudes. L'attribution de traits propres à un groupe différent de celui de l'attributeur, la différence entre l'homogénéité perçue de ce groupe et la variété dans son propre groupe sont étudiées, et rapportées à des processus de catégorisation différents. Dans les sections suivantes seront évoqués d'autres thèmes à l'interface des sciences cognitives et des sciences sociales.

(c) *Vers une théorie des cultures en forme de psycho-anthropologie évolutionniste.* Les sciences cognitives (toutes disciplines conjointes), la paléontologie, l'anthropologie, l'archéologie, l'éthologie, et enfin la théorie de l'évolution semblent donner naissance, depuis une dizaine d'années, à une nouvelle approche des sciences sociales, comprenant en son centre un projet de science naturelle de la culture. Les développements récents de la psychologie et des neurosciences cognitives, en particulier la domanialité, l'hypothèse darwinienne et les données sur l'évolution biologique et culturelle de l'espèce se soutiennent mutuellement.

Les cultures humaines ont beaucoup en commun avec les espèces vivantes: elles sont soumises à des processus évolutifs; elles sont des phénomènes collectifs soumis à deux ordres de contraintes ("verticales", en provenance des individus, "horizontales", en provenance des populations) et exerçant dans ces deux directions des contraintes réciproques; elles présentent une variabilité apparemment illimitée, qui laisse néanmoins prise à des invariants; elles ont émergé d'un état dans lequel elles n'existaient pas; elles présentent des traits adaptatifs, et des traits antiadaptatifs. Pour autant, espèces vivantes et cultures sont des entités très différentes. Il ne peut donc s'agir de plaquer sur les cultures des concepts et des schèmes explicatifs de la théorie de l'évolution, mais de s'inspirer de cette théorie pour dégager les conditions d'une "pensée populationnelle" (le *population thinking* qu'Ernst Mayr a théorisé et popularisé en biologie) adaptée à l'étude des cultures.

Ce néodarwinisme culturel se distingue, d'une part, de tentatives antérieures, remontant à l'époque de Darwin, pour appliquer aux phénomènes sociaux des schémas évolutionnistes généraux sans articulation avec les mécanismes matériels; et d'autre part, de l'application de la théorie néodarwinienne proprement dite à la détermination des contraintes fonctionnelles au niveau individuel. La théorie de l'évolution joue d'une part un rôle heuristique au niveau populationnel, d'autre part un rôle directement applicatif, dans le cadre de la psychologie évolutionniste. Telle est la formule de la nouvelle approche proposée pour naturaliser les sciences sociales sans les réduire à la biologie ou à la psychologie, une approche dont John Tooby et Leda Cosmides ont publié le manifeste en 1992 en tête du recueil *The Adapted Mind*. Ce que la dernière décennie a apporté, c'est une floraison de travaux s'inscrivant dans ce programme, plus particulièrement dans le domaine de la culture. L'extrême sensibilité du sujet conduit évidemment à des polémiques, et il est nécessaire de ne pas perdre de vue les deux composantes du programme : la composante naturaliste au sens strict, traditionnel du terme, porte sur les mécanismes, individuels et extra-individuels, du traitement de l'information; elle n'entre pas en contradiction avec la seconde composante, à savoir le projet, également traditionnel, de rendre compte de la variété des cultures et de fournir une explication compréhensive de leurs contenus spécifiques, fondée sur un apprentissage et une transmission. Mais cette explication doit néanmoins s'intégrer dans le cadre naturaliste, ce que le double recours à l'évolution (évolution biologique et évolution culturelle) semble rendre au moins concevable.

Cette intégration n'est pas conçue tout à fait de la même manière parmi les différents chercheurs, au reste encore peu nombreux, qui se réclament de cette nouvelle psycho-anthropologie. Si aucun ne se veut déterrer la hache de guerre entre "nature" et "culture", certains sont à leur façon plus culturalistes que d'autres : pour Robert Boyd et Peter Richerson, dont *La nature de la culture et des cultures*, attendu depuis longtemps, est sur le point de paraître, la culture est une capacité biologique essentielle de l'espèce, reposant largement sur une capacité d'imitation singulièrement développée (elle-même conditionnée par l'aptitude à l'attention conjointe). La psychologie cognitive joue un rôle certes essentiel, mais moins structurant que dans le courant domanialiste, puisque

l'accent est mis sur la capacité transdomaniale de l'imitation et, de ce fait, sur les contenus tels qu'ils sont généralement appréhendés par les anthropologues.

Il ne s'agit pas non plus d'un retour de la sociobiologie (dont le rejet violent, particulièrement dans notre pays, appellerait d'ailleurs un réexamen). Car l'on ne suppose pas, comme le faisait la sociobiologie, que les comportements culturels humains sont adaptatifs, pour la double raison que la théorie de l'évolution n'implique pas l'adaptation de tout ce qui existe, pas plus que l'adaptation ne s'explique directement par la sélection naturelle; et que l'environnement dans lequel les forces de sélection se sont exercées pour façonner l'équipement cognitif humain est celui des chasseurs-cueilleurs du pléistocène, un environnement qui n'est pas le nôtre et dont nous ne pouvons qu'inférer les caractéristiques.

Parmi les thèmes les plus centraux de ce courant de recherche, outre la nature et l'émergence de la culture, la part respective de la variabilité et de l'universalité dans les cultures, et l'évolution culturelle, que je viens d'évoquer, figurent toutes les grandes questions de l'anthropologie, en particulier celle des mythes et des croyances religieuses, croyances contre-intuitives ou apparemment irrationnelles. Un autre problème-clé est celui de l'altruisme et de la coopération : comment peuvent-ils émerger, en contradiction apparente avec l'adaptativité massive de l'égoïsme? et, plus difficile encore, comment peuvent-ils se maintenir, face à l'avantage immédiat de la resquille? Les réponses avancées reposent, pour certaines, sur des mécanismes purement évolutifs (sélection de la parentèle, un mécanisme proposé par D. Hamilton en 1964; sélection de groupe, une théorie discréditée mais remise en selle récemment par E. Sober et D. Wilson), pour d'autres, sur des considérations de théorie économique, on va le voir dans un instant.

(d) *Cognition animale et approches comparatives; la question du langage.* L'intérêt pour la "psychologie" animale (on a longtemps plus volontiers parlé d'intelligence animale) est aussi vieux que la psychologie elle-même, et que sa discipline-mère, la philosophie; deux de ses courants modernes, la *Gestalt* et le béhaviorisme, ont consacré une énergie considérable à lui appliquer leurs concepts et leurs méthodes empiriques (ce qui ne va pas de soi). Les sciences cognitives rejoignent cette tradition, et accordent à l'animal, pour plusieurs raisons, une importance croissante.

Sur le plan le plus général, la *cognition* animale fournit à la fois un sujet d'étude fascinant, un formidable atelier méthodologique, la base d'un comparatisme cognitif, et la source irremplaçable de modèles. Comme son prédécesseur immédiat, le béhaviorisme, la psychologie cognitive a pour obligation et pour pierre de touche la capacité de s'appliquer aussi bien à l'animal "non humain" qu'à l'homme : elle ne pouvait donc en différer longtemps l'épreuve. Mais elle l'aborde dans une perspective différente. La tradition philosophique retient chez l'animal, comme caractéristique objective essentielle, son comportement, et chez l'homme son esprit (dont le langage témoigne avec une abondance de détails). Le naturalisme, au sens épistémologique large comme au sens étroit associé aux naturalistes et à Darwin, postule une continuité entre animal et homme. Le béhaviorisme exprime cette continuité en alignant l'explanandum relatif au second sur le premier : pour l'homme désormais, comme traditionnellement pour l'animal, c'est du comportement qu'il faut se préoccuper exclusivement. La psychologie cognitive, sans renverser complètement les choses, étend à l'animal le privilège de l'esprit, tout en conservant, dans son cas comme dans celui de l'homme, le comportement comme trace essentielle.

Les rapports entre psychologie scientifique humaine et éthologie (l'étude scientifique du comportement animal) sont complexes, et mettent en jeu des oppositions méthodologiques et

ontologiques largement préfigurées par les débats entre Edward C. Tolman et ses critiques dans les années 1930–40. Les deux principales portent sur la méthodologie : observation en milieu naturel (problème de la “validité écologique”) ou expérimentation contrôlée en laboratoire (pour surmonter les limites des “preuves anecdotiques”), d’une part, statut réaliste ou bien purement heuristique et provisoire des descriptions et des explications “mentalistes”, subjectives ou intentionnelles, d’autre part. L’éthologie cognitive se sépare de l’éthologie traditionnelle et de la psychologie comparative en conservant de la première le souci de l’observation *in situ*, généralement rejeté par la seconde, et de cette dernière le réalisme intentionnel, rejeté par la première.

Or, pour explorer l’espace conceptuel du vocabulaire intentionnel, l’apport de la méthode comparative, si délicate qu’elle soit à déployer, est irremplaçable. Elle permet de séparer par la pensée des capacités qui se trouvent, apparemment, présentes et indissociables chez l’homme, de rechercher pour chacune des preuves de leur existence, ou du moins des raisons de les admettre dans l’ontologie scientifique, et de mettre au jour les relations, contingentes ou nécessaires, qu’elles entretiennent. C’est aussi ce qu’on attend de la neuropsychologie (avec des obstacles théoriques et pratiques différents), mais l’espèce animale exhibe une fonctionnalité, une adaptation à son habitat, une intégration de ses capacités, qui rendent praticables les méthodes éprouvées de la biologie évolutionniste. La question des liens entre “phénoménologie” et description scientifique, entre “monde vécu” ou sujet et manifestation publique ou objet (ou encore entre les perspectives de la première et de la troisième personne) prend un relief particulier dès qu’on l’élargit aux créatures non humaines : c’est un thème bien connu du public, c’est aussi le cœur de l’éthologie “des profondeurs” dont von Uexküll peut être considéré comme le pionnier, et Donald Griffin le maître contemporain.

Plus généralement encore, si, de leur côté, les sciences cognitives s’intéressent désormais tellement aux études sur l’animal, c’est parce qu’elles n’envisagent plus la cognition de l’homme adulte indépendamment de sa genèse : *phylogenèse*, *ontogenèse*, “*neurogenèse*”. Or l’animal fournit, en ce qui concerne la première perspective, des points de repère indispensables sur l’arbre de l’évolution; en ce qui concerne la deuxième, des comparaisons entre les états initiaux (génétiquement déterminés) d’espèces différentes (en particulier leurs “instincts pour apprendre”) et les trajectoires menant, par développement et apprentissage, aux états finaux de ces espèces; en ce qui concerne la troisième, des comparaisons entre les systèmes nerveux centraux, génératrices d’inférences, à partir de correspondances attestées chez l’animal (entre images et tissus ou processus, entre structures et fonctions, entre lésions, actions pharmacologiques ou privations et déficits,...), sur des correspondances analogues chez l’homme, impossibles à établir directement pour des raisons éthiques. Curieusement, en retour, le progrès des connaissances entraîne une prise de conscience de la “personne” animale et de ses droits, susceptible à terme de rendre plus délicat le recours au modèle animal.

Une dernière grande raison de l’intérêt pour les espèces animales est qu’elles exhibent une considérable variété de formes spécifiques de “cognition sociale”. La recherche d’explications “distales” (évolutionnistes) des différentes structures et relations sociales non humaines ouvre des horizons pour l’étude des bases cognitives de la socialité humaine. Altruisme, stratification sociale, différenciation sexuelle et choix des partenaires, autant de questions que l’éthologie cognitive renouvelle. La communication en est une autre, peut-être la plus importante dans la mesure où elle conduit directement à la question “maudite” de l’origine du langage et des capacités prélinguistiques d’espèces animales. Cette question est à son tour étroitement liée à celle des liens entre pensée et langage et de la nature et de la possibilité d’une pensée sans langage. Difficilement décidables

lorsqu'on les confine à l'homme, elles se laissent attaquer dans la perspective large du nonhumain (comme, du reste, du non-normal).

(e) *Economie : les courants expérimentaux et cognitifs se rejoignent.* L'idée que les propriétés du modèle d'agent, dit *Homo œconomicus*, sur lequel repose la théorie économique, doivent être en rapport avec celles des agents économiques réels, sans être universellement acceptée par les économistes, est néanmoins ancienne : Hayek la défendit et l'illustra avec les moyens que lui fournissait la psychologie de son époque. Dès les années 1950, elle inspire des travaux de psychologie sociale visant à mieux cerner ces rapports. Vers la fin de cette décennie naît l'économie expérimentale, aux Etats-Unis et indépendamment en Europe. Les courants psychologique et économique resteront longtemps séparés, malgré la jonction qu'on aurait pu croire décisive opérée par Herbert Simon, qui fut à la fois, à la même époque, l'un des pères fondateurs des sciences cognitives (avant la lettre) et prix Nobel de sciences économiques (en 1976). La notion de rationalité limitée qu'il propose a pour raison d'être de replacer l'agent économique dans l'univers réel sans le rendre inapte à tenir son rôle dans la théorie classique de la décision, convenablement ajustée. *Homo œconomicus simonicus* est limité dans ses capacités de calcul, et applique par conséquent des heuristiques qui le conduisent en probabilité à des décisions sub-optimales dans l'absolu (celles que prendrait à sa place *Homo œconomicus classicus*), mais d'un coût épistémique réaliste. Il demeure cependant un être abstrait, assimilable à un algorithme (sa principale vertu aux yeux de son créateur).

Cette prise en compte de la matérialité psychologique de l'agent relève d'un "cognitivism faible", qui consiste à substituer à un modèle très idéalisé de l'agent un modèle qui l'est un peu moins, mais reste cependant très éloigné d'*Homo sapiens*, organisme existant, bricolage hypercomplexe de l'évolution et objet véritable des sciences cognitives. Un pas vers davantage de réalisme (au double sens épistémologique : remplacer les hypothèses instrumentales par des faits empiriques, et psychologique : remplacer l'homme-algorithme par un système organique, différencié, modulaire) est accompli par le courant de recherches des heuristiques et biais (*heuristics and biases*) animé depuis le milieu des années 1970 par Amos Tversky (prématurément disparu) et Daniel Kahneman (qui participa à la décade de Cerisy dont est issu ce volume, et obtint, après la mort de Tversky, le prix Nobel de sciences économiques pour ces travaux) — cf. chap. 11. Alors que Simon mettait l'accent sur les capacités de calcul de l'agent, leur assignant une finitude abstraite, Tversky et Kahneman s'attachent à montrer expérimentalement les imperfections graves de son système d'évaluation des probabilités. Les expériences, très robustes, qu'ils mènent les conduisent, et de nombreux philosophes avec eux, à des conclusions pessimistes sur la rationalité humaine. Gerd Gigerenzer conteste la validité de ces conclusions, et même si l'issue de la controverse reste incertaine, on observe aujourd'hui un rééquilibrage de la psychologie sociale, moins exclusivement tournée vers la mise au jour des défauts de l'agent économique et social, et cherchant à cerner également ses qualités. Certains philosophes jugent que la rationalité mise en défaut par les "illusions cognitives" de Tversky et Kahneman n'a au plus qu'un rapport indirect avec la rationalité authentique. Tout cela contribue à accélérer la convergence entre sciences normatives et sciences descriptives de la rationalité, d'une part parce que la concurrence des hypothèses appelle un recours beaucoup plus approfondi aux concepts et théories des disciplines pertinentes, et un renforcement graduel du cognitivism faible des débuts; d'autre part parce qu'en déplaçant son attention des défauts vers les qualités du système cognitif humain, la "psycho-économie" et la psychologie sociale plus généralement s'arriment plus fortement à la psychologie évolutionniste et aux neurosciences.

Un troisième axe de pénétration du nouveau réalisme en théorie économique porte sur les préférences, que l'on cesse de prendre pour des faits bruts, et qu'on cherche à articuler d'un côté avec une conception des motivations plus large que la famille des intérêts bien compris, de l'autre avec une prise en compte théorique des émotions (qui ne sont plus un simple bruit parasitant le système de décision). Le premier volet est abordé avec les méthodes de l'économie expérimentale, qui met à l'épreuve des hypothèses étonnamment banales sur l'envie de rendre la pareille (en bien ou en mal), le besoin d'approbation ou le désintéressement (monétaire) comme facteurs essentiels de choix. Le second volet relève davantage d'une élucidation conceptuelle, menée d'une part à partir du fonds de la philosophie morale, de la littérature et de l'histoire, d'autre part en liaison directe avec la réhabilitation des émotions comme thème pour la psychologie cognitive, les neurosciences et la philosophie de l'esprit (on y vient en (f) *infra*).

Mais les apports les plus novateurs concernent le rapport à autrui, appréhendé comme une fonction, soit dans son plein développement soit dans son acquisition.

La fonction considérée consiste en ceci. A tout moment l'agent détermine sa réaction à l'action d'autrui et l'on peut, dans un premier temps, étudier les stratégies qu'il applique, ce qui est l'objet de la théorie des jeux. Sans quitter ce cadre, on peut examiner deux phénomènes ignorés par les approches classiques. Le premier est l'altruisme : l'étude expérimentale confirme l'intuition selon laquelle certains comportements de certains agents dans certaines circonstances revêtent au moins l'apparence de l'altruisme, et l'on peut identifier, par l'analyse conceptuelle et par l'expérimentation, différentes formes et degrés de l'altruisme, en distinguant notamment différentes situations de coopération et de réciprocité impliquant une pluralité d'agents. Le second est l'anticipation, c'est-à-dire la prédiction de l'action de l'autre à partir de ses intentions supposées, ce qui implique déjà le passage d'une stratégie "béhavioriste" à une stratégie "mentaliste", et se complexifie dès qu'on considère que la pensée de l'autre suit le même chemin, en sorte que les anticipations de chacun se fondent sur les anticipations de son vis-à-vis.

L'étape suivante consiste à introduire le temps de l'apprentissage individuel : une stratégie peut être apprise, soit par l'expérience, soit par la transmission pédagogique ou culturelle. Les travaux les plus précis portent sur la première forme : une suite de parties du même jeu avec le même partenaire conduit à la stabilisation d'une stratégie particulière (sous certaines conditions).

Enfin, l'apprentissage peut être recherché du côté de la phylogenèse et non de l'ontogenèse, ce qui fait passer au mode de pensée "populationnel", selon l'expression d'Ernst Mayr. Autrui est alors, non pas un ou plusieurs partenaires et concurrents, mais une population d'agents, porteurs de préférences, de normes et de stratégies, qui sont autant de caractères "phénotypiques" répandus en proportions différentes, et variables dans le temps. Cette "génétique des populations" se combine à des hypothèses sélectionnistes au sein d'un analogue de la « nouvelle synthèse » néodarwinienne dont l'objectif est d'expliquer la dynamique des normes, préférences et stratégies, dynamique comportant des disparitions et des apparitions d'espèces, et des stases ou équilibres caractérisés par certaines proportions des traits étudiés. Une dimension supplémentaire, sans homologue dans la théorie évolutionniste, est celle des institutions, vues comme systèmes de coordination émergents agissant non pas, tels la nature ou la niche écologique, directement sur la survie matérielle des agents, mais indirectement sur leurs représentations internes, lesquelles agissent en retour sur les institutions.

(f) *Emotions et affects. La cognition chaude est de retour. A vaincre sans péril on triomphe sans gloire : il n'est pas bien difficile de convaincre notre époque narcissique et sentimentale que les*

émotions jouent un rôle déterminant dans notre vie mentale, dont elles “accompagnent”, “colorent”, et peut-être “dirigent” ou du moins co-dirigent la plupart des moments, sinon tous, et que les sciences cognitives ont eu tort, en invoquant trop longtemps une distinction inanalysée et provisoire entre cognition et émotion, de les négliger. Qu'on ajoute, avec toute la solennité scientifique nécessaire, que les émotions s'inscrivent de telle et telle manière dans les processus cérébraux et qu'elles ont une fonction évolutionnaire, et le succès est garanti.

Mais ces effets médiatiques ne sont que le reflet d'une activité scientifique et philosophique bien réelle, qui sans avoir jamais complètement cessé, depuis Darwin et William James, a connu un développement considérable au cours des années 1990, et constitue indubitablement un trait caractéristique du *Zeitgeist* des sciences cognitives actuelles.

Les émotions occupent en effet une position stratégique qu'elles doivent tant à leurs énigmatiques propriétés qu'à la conjoncture présente. Historiquement, les émotions viennent à point nommé renforcer le retour du balancier de la “cognition froide” vers la “cognition chaude”, de l'intellectualisme vers l'incarnation, de l'hyper-rationalisme vers une forme d'intuitionnisme, du détachement comme mode fondamental vers l'implication et le “situationnisme”, de l'étude *in vitro* vers l'observation *in vivo*, etc. Au sein du domaine, les émotions, comme nul autre sujet, mobilisent autant les neurosciences que la philosophie, la psychologie, l'éthologie et l'anthropologie. Aux frontières du domaine, elles établissent ou rétablissent de multiples connexions : avec la psychanalyse et la psychologie sociale; avec l'économie; avec la théorie de l'évolution; avec la philosophie des sciences, avec la philosophie morale et la philosophie politique; avec la phénoménologie; avec les théories de la culture, de la littérature et des arts.

Objet insaisissable, versatile, aux traits contradictoires : les émotions donnent du fil à retordre aux philosophes, dont l'une des principales fonctions est de faire de l'ordre dans le fouillis des mots et des concepts. Ils s'emploient à établir une taxinomie des émotions, et se divisent entre ceux qui estiment qu'il existe un taxon maximal, les émotions comme espèce naturelle, et ceux qui estiment que seules certaines d'entre elles forment ensemble une catégorie naturelle : ce sont les émotions de base ou “programmes affectifs” liés aux fonctions éthobiologiques fondamentales, transculturelles, inscrites dans le développement de l'enfant normal, exprimées spécifiquement sur le visage, etc. On compte, selon les auteurs, six à huit émotions de base : tristesse, dégoût, colère, joie, peur, surprise, attente, acquiescement. Les liens entre les différentes émotions, entre émotions et sentiments, entre émotions et désirs, entre émotions et croyances sont autant de problèmes épineux. L'un des traits les plus singuliers des émotions est leur dimension normative : toute épisode émotif s'accompagne d'une évaluation, d'une appréciation (*appraisal* en anglais) portée sur l'objet de l'émotion (à supposer, ce qui ne va pas de soi, que toute émotion possède l'intentionnalité), sans nécessairement s'y réduire. D'autre part, les émotions sont, directement et indirectement, étroitement liées à la personnalité, au soi, à la subjectivité, à la conscience. Ainsi les émotions sont-elles étroitement liées à la vie morale. Cette configuration de propriétés accreditte l'idée que les émotions sont des entités mentales *sui generis*.

Mais si c'est le cas, il faut s'attendre que leur “base neurale”, leur contrepartie matérielle, sera elle aussi d'un genre particulier, ou encore que le rapport entre les émotions comme entité psychologique, phénoménologique ou comportementale et les processus cérébraux qui les “soutiennent” sera lui-même d'un genre particulier. Les neurobiologistes ont là-dessus beaucoup de choses à dire. Ils tendent même à considérer qu'il leur revient de dire ce qui mérite, à ce sujet, d'être dit. L'une des raisons pour lesquelles on leur prête attention porte un nom : Phineas Gage, en passe de devenir le Victor de notre époque, comme son “père scientifique”, Antonio Damasio, est notre Dr Itard

(le médecin de l'enfant-loup de l'Aveyron), un Dr Itard posthume s'entend. En perdant à l'âge de 25 ans, suite à un peu banal accident du travail survenu en 1848, une bonne partie de son cortex orbito-frontal, Gage fut privé non seulement de la capacité d'éprouver des émotions normales, mais également celle de mener une existence professionnelle et sociale, alors qu'il était, avant l'accident, un parangon des vertus sociales, époux, citoyen et contremaître modèle. L'hypothèse de Damasio, fondée sur les témoignages (nombreux, car Gage eut de son vivant son véritable Dr Itard, le Dr Harlow), et sur l'observation de cas contemporains, est double : la lésion de Gage préservait ses fonctions purement cognitives, mais lui ôtait à la fois la capacité de prendre les décisions appropriées, d'avoir des relations sociales normales, et d'éprouver des émotions : celles-ci constituent donc un volet crucial de la rationalité, et parce que pour l'être humain, animal social, la rationalité consiste essentiellement à naviguer parmi ses congénères en optimisant son propre sort, les émotions sont le pilote de notre vie sociale; elles biaisent le calcul d'utilité en engendrant des "marqueurs somatiques", assez semblables aux réflexes pavloviens, mais plus flexibles, et comme eux hautement adaptatifs. Privée de chaleur, la cognition n'est pas seulement froide : elle est inopérante. Ainsi, les dichotomies traditionnelles cognition/émotion, cœur/raison, corps/esprit, sont-elles mises à mal - pauvre Descartes! triste tradition intellectualiste!

Si j'ai rappelé l'hypothèse de Damasio, bien connue aujourd'hui, c'est parce qu'elle réunit la plupart des ingrédients de la fascination qu'exerce le thème des émotions sur bon nombre de philosophes, professionnels ou non. *Primo*, les émotions, considérées depuis l'Antiquité comme les antagonistes de la raison, se voient désormais attribuer un rôle essentiel aux côtés du *calculus ratiocinator*. Elles ont une *fonction cognitive*, au lieu d'être, peut-être, un mal nécessaire (un compromis imaginé par la Création pour assurer la survie de l'animal sans trop compromettre la vie de l'esprit). *Secundo*, pour cette raison et parce qu'elles semblent être inextricablement somatiques et psychiques, les émotions brouillent complètement l'ontologie du mental. *Tertio*, elles sont au fondement de notre rapport à autrui, ce qui justifie notamment leur place dans la seconde partie du présent panorama, malgré leur proximité au thème de la conscience, placé dans la première, qui est l'autre terrain de bataille où neurologues-médecins et philosophes se disputent l'honneur de mettre en lambeaux de ce qu'ils croient être la doctrine cartésienne. Enfin, elles sont adaptatives, et alimentent la réaction optimiste aux tenants de l'irrationalité foncière de l'homme, contribuant aux efforts déjà mentionnés pour donner consistance à une "vision positive" de la nature humaine.

Il demeure, comme le souligne fortement Anthony Marcel, que nous sommes encore bien loin de posséder une vue cohérente de l'expérience émotionnelle, de ce que c'est que d'éprouver une émotion, et que nous sommes à cet endroit dans des perplexités semblables à celles que soulèvent l'action et la conscience, peut-être pour les mêmes raisons. Une révision déchirante s'annonce, au terme de laquelle la conception scientifique et la conception ordinaire du mental pourraient se retrouver définitivement éloignées l'une de l'autre. C'est du reste ce dont ne cesse de nous avertir Dennett depuis une trentaine d'années.

C. LES « TECHNOLOGIES CONVERGENTES » : « NBIC » ET LES SCIENCES COGNITIVES APPLIQUÉES

En produisant de minuscules machines motrices, cognitives et communicatives, susceptibles de s'ajouter, d'un côté, aux cellules et en particulier au système nerveux, central et périphérique, d'un individu, de l'autre, aux réseaux informationnels et énergétiques planétaires, et, entre les deux, aux

corps vivants (organes et muscles) et aux objets et appareils environnants, les nanosciences modifieraient la condition humaine, dans les vingt prochaines années, plus radicalement qu'on l'a jamais fait ou imaginé. Dans un avenir prévisible, en tout cas, des changements considérables, dont les prémices apparaissent déjà, affecteraient quasiment tous les secteurs de l'activité humaine, de la médecine aux loisirs, de la politique à l'éducation, de la défense et la police à l'industrie et la science.

Tel est l'avenir qu'annonce et décline le retentissant rapport d'un *think tank* indépendant, le WTEC (World Technology Evaluation Center), rendu public en juillet 2002 par la National Science Foundation (NSF) et le Department of Commerce américain. Ces bouleversements résulteraient de la synergie entre les quatre « technologies » maîtresses du XXI^e siècle, N pour *nano*, B pour *bio*, I pour *info*, C pour *cogno*, d'où le signe NBIC.

On évoque parfois à ce propos une révolution épistémologique. Certains font grand cas d'une mutation ontologique et anthropologique résultant de la possibilité même de construire des systèmes artificiels non plus à l'échelle de la physique macroscopique et selon les méthodes classiques d'assemblage, mais à celle des atomes et des molécules, comme dans les réactions chimiques, tout en conservant, du paradigme de l'ingénieur mécanicien, l'universalité de la combinatoire pièce par pièce en "milieu sec" (rendant possible un découplage intégral et une discrétisation des opérations élémentaires, contrairement au milieu fluide dans lequel opèrent les processus chimiques). D'autres voient dans la convergence des quatre "technologies" non pas seulement un événement historique probable ou certain, souhaitable ou menaçant, mais un appel à abattre les frontières disciplinaires et à créer une *scienza nuova* unifiée, une nouvelle philosophie naturelle, reposant sur une synthèse d'un nouveau genre baptisée "consilience". Certains estiment que l'échelle nanométrique offre la possibilité de réaliser enfin pleinement la vision des fondateurs de la cybernétique, et pose les problèmes scientifiques qu'ils avaient formulés, ceux de l'auto-organisation, de l'auto-réplication et de la complexité. On évoque aussi l'avènement d'un émergentisme absolu, à partir des composants les plus élémentaires : une victoire finale, tant conceptuelle que pratique, du *bottom-up* sur le *top-down*. On pose une équivalence entre les quatre types de composants de base de la nouvelle techno-nature : les atomes du physicien et du nano-ingénieur, les neurones du neurobiologiste, les gènes du biologiste moléculaire, les bits de l'informaticien, d'où la désignation plaisante, par l'ETC, un groupe canadien, de "Little B.A.N.G." (B pour bit, A pour atome, N pour neurone, G pour gène); on y ajoute parfois les memes [sans circonflexe] de Richard Dawkins, unité élémentaire de culture. *Last but not least*, le passage au "nano", qui, on va le voir, sous-tend l'ensemble du mouvement, est parfois considéré comme l'ultime frontière à ne pas franchir, conférant au Prométhée qui s'y risquerait une puissance jamais atteinte, grosse d'une menace mortelle; selon l'ETC, "*size matters*", la taille fait une différence, et il faut imposer un moratoire à ces recherches (comme celui que s'imposèrent il y a trente ans, à Asilomar, les biologistes moléculaires auxquels les manipulations génétiques semblaient promettre à la fois un avenir radieux et le risque d'une catastrophe d'une gravité sans précédent).

Je laisserai cependant de côté ces spéculations, et ne considérerai que ce qui, dans NBIC, est de nature à influencer sur le devenir des sciences cognitives.

(a) *Les quatre protagonistes et la question de la technologie.* Les sommets du "tétraèdre NBIC" ne se laissent pas définir aisément, mais en gros, N désigne à la fois la "nanoscience" et la "nanotechnologie": la première consiste en l'étude des phénomènes à l'échelle nanométrique (en pratique, aux échelles atomique, moléculaire et macromoléculaire, soit de 1 à quelque 100 nanomètres, un nanomètre étant un millième de micron, soit 10^{-9} m) et de la manipulation de

matériaux à cette échelle; la seconde rassemble les procédés de fabrication de structures et de systèmes reposant sur la manipulation d'objets nanométriques. I réunit l'informatique et les sciences et techniques de l'information et de la communication (STIC). Par rapport aux deux autres, les sommets N et I sont "formels", dans la mesure où ils ne sont liés à aucun domaine naturel particulier. B désigne les biotechnologies, mais également la biologie théorique, conçue comme l'étude des conditions générales d'existence des systèmes vivants, les neurosciences, et les sciences médicales. C a la singularité de renvoyer explicitement aux *sciences* cognitives, et non à une technologie encore à imaginer, à moins qu'elle ne soit pensée comme incluse dans les zones d'influence de la biologie et de l'informatique; ce qui importe peu, au fond, si l'on prend au sérieux l'idée même de la convergence.

Si les sciences cognitives sont dénuées d'un volet technologique propre, les trois autres disciplines dont on annonce la convergence ne sont pas de pures technologies, mais comportent chacune, on le voit, une dimension scientifique théorique. La coloration spécifiquement technologique apparaît dans le processus de convergence lui-même, du moins dans la perspective adoptée dans le rapport du WTEC. Cette perspective s'explique à son tour par l'objectif immédiat des auteurs, commanditaires et promoteurs du rapport, à savoir convaincre les parlementaires américains de voter des crédits considérables pour développer les recherches dans ce nouveau domaine. Elle reflète aussi l'état d'esprit d'une certaine Amérique, agonistique (la compétition économique et la suprématie militaire) et mélioriste ("améliorer le fonctionnement [*performance*] de l'homme", tel est l'objectif figurant dans le titre même du fameux rapport). Enfin et surtout, elle correspond à une véritable question, celle de l'existence, individuelle et collective, qui sera la nôtre au XXI^e siècle. Ici l'analyse critique du rapport lui-même, des contingences de sa production, de ses extrapolations hasardeuses, de son idéologie implicite doit céder le pas à une réflexion réaliste mais pas myope, techniquement informée mais s'élargissant en une méditation politique, historique et anthropologique. Ce n'est évidemment pas le lieu d'en esquisser les contours, mais il est significatif qu'au moment où j'écris ces lignes, un groupe réuni par le Directorate de la recherche à Bruxelles s'attelle à cette tâche.

(b) *Les liens entre C et N, B, I.* Les sciences cognitives sont directement reliées aux sommets I et B. Les interactions avec I sont ce qui porte encore souvent le nom d'intelligence artificielle (IA), tout en recouvrant quelque chose d'assez différent de l'IA des origines, ainsi que l'ergonomie cognitive comprise désormais comme science des interfaces hommes-machine ou encore des "usages", les nouvelles technologies éducatives, commerciales, économiques, industrielles, décisionnelles... l'"intelligence ambiante" (*ambient intelligence*). Les interactions avec B relèvent des neurosciences, avec un volet proprement cognitif, un volet cellulaire et moléculaire, un volet pharmacologique et médical; elles incluent également la vie artificielle, rattachée à la nouvelle IA. Rien de bien nouveau, du point de vue des sciences cognitives, mais indicatif ou annonciateur d'une prise de conscience au niveau politique, d'un élargissement de perspective, et d'un saut quantitatif.

En combinant B et I, on obtient des interactions avec C qui, sans être inédites, passent aujourd'hui, ou passeront bientôt, des prémices à la pratique courante. Du côté de la science : la neuroinformatique, un concept aux contours encore variables sous lequel on peut subsumer la modélisation des neurones et des structures cérébrales "biologiquement plausibles", l'étude du "codage neural", ou traitement de l'information dans le tissu neural, le traitement des données *dans* les techniques d'imagerie cérébrale, structurelle et fonctionnelle, et les méthodes d'analyse des *résultats* obtenus; la psychophysique "neuroscientifique", la neuro-robotique... Du côté des applications : l'"*e-health*", ou médecine par internet (dont une préfiguration rudimentaire est le

système-expert d'aide aux diabétiques d'il y a 20 ans), et plus largement la "médecine ambiante" (*ambient healthcare*). Les prothèses sensorielles et motrices et les systèmes de commande directe à partir du cerveau d'effecteurs et d'appareils se situent à mi-chemin entre ces deux domaines. Remarquons au passage l'ambiguïté du "I", qui désigne tantôt le concept théorique d'information, consubstantiel au projet des sciences cognitives, tantôt le concept ordinaire.

Quid de N? N est partout, ou pour jouer sur le mot-fétiche du moment, il est "ambiant". Il tient une multitude de rôles hétérogènes. Ainsi, toute connexion matérielle avec des cellules, nerveuses ou autres, toute intervention directe sur les macromolécules, se fait, par définition, à l'échelle nanométrique et requiert donc nanoscience et nanotechnologies. L'ultraminaturisation des processeurs tend vers l'échelle nanométrique, et induit un saut quantitatif en informatique, avec les conséquences que cela implique aussi bien pour les outils que pour les applications des sciences cognitives. Les relais matériels de l'"intelligence ambiante" dépendent à la fois de la "nanométrisation" des composants, des dispositifs de visualisation et des systèmes de communication, lesquels seront tributaires des nanotubes et autres composants issus des nanotechnologies.

(c) *Le retour de l'IA prométhéenne*. Qui a connu l'ascension et la chute de l'IA classique éprouve en lisant certains manifestes des technologies convergentes un fort sentiment de déjà-vu. Il retrouve la liste sans fin des réalisations, tous domaines confondus, que l'informatique intelligente, parvenue au stade ultime de l'"intelligence ambiante", nous promet à moyen terme. Il retrouve les craintes que soulève la perspective des progrès à venir (il y a trente ans, nous étions invités à nous inquiéter de ce qu'il adviendrait de la politique le jour, proche, où le premier ministre serait un ordinateur "intelligent").

Il retrouve surtout l'erreur princeps de l'IA "prométhéenne", celle de croire que l'on peut simuler ou amplifier l'esprit, et ses diverses facultés et productions, individuelles et collectives, en prenant pour modèle, au sens antique, artistique du terme, les conceptualisations qu'en livre la connaissance commune. Nous savons, semblent penser nos nouveaux artificialistes, ce que c'est que voir, toucher, apprendre, mémoriser, raisonner, communiquer, s'émouvoir, décider, éprouver de l'empathie, planifier une action motrice, coopérer, etc., et nous avons une idée assez claire des liens qui existent entre ces phénomènes. La difficulté, selon eux, consiste à traduire cette connaissance en des algorithmes et des processus de capture et d'émission d'information, exécutables matériellement, dans de bonnes conditions de fiabilité, par des dispositifs physiques et des systèmes computationnels réalisables, sur le plan technologique, ergonomique, économique, etc.

Tout se passe comme si une leçon avait été oubliée, lors du passage de relais d'une génération à la suivante : le pari de l'IA classique était qu'en simulant les processus intelligents, appréhendés à partir de leurs *résultats*, caractérisés dans les termes de la connaissance commune, on parviendrait à la théorie explicative de ces processus : c'était toute l'idée des "programmes comme théories" (*programs as theories*). En élaborant, par approximations et généralisations successives, des programmes conduisant aux "performances" observées, on obtiendrait la "compétence" correspondante. Ce pari a été perdu : on ne peut faire l'économie d'une théorie scientifique traditionnelle, c'est-à-dire appuyée sur des concepts abstraits spécifiques au domaine visé, et visant à cerner, en un processus de conjectures et d'expériences alternées, la "compétence" psychologique susceptible d'expliquer les "performances" observées. C'est une fois cette compétence caractérisée, fût-ce provisoirement, qu'on peut entreprendre de la reproduire, dans ses traits essentiels, sous forme de programmes.

L'IA d'aujourd'hui, assagie, l'a bien compris, qui se divise en spécialités telles que logique appliquée, vision artificielle, traitement automatique des langues, etc., et conçoit son rôle comme celui d'un passeur entre disciplines fondamentales et informatique. Mais l'effet d'entraînement des technologies convergentes semble avoir recréé les conditions sociologiques et intellectuelles d'une nouvelle illusion de toute-puissance.

(d) *Science et technologie*. Il ne faut pas tomber dans l'excès inverse : la technique, prise au sens le plus large, ne dépend qu'en partie de la science. Elle dégage selon ses besoins des ressources épistémiques dont il n'apparaît souvent qu'après coup qu'elles sont des "applications" d'une science venue plus tard. L'homme fabrique des outils et des demeures merveilleusement adaptés à ses capacités motrices et visuelles depuis des millénaires, alors que sa connaissance scientifique de ces capacités ne se dégage que depuis peu. Les machines complexes et les appareils sont apparus plus récemment, mais précèdent encore de beaucoup la naissance des sciences cognitives. Même l'ergonomie, science de l'homme au travail, particulièrement au travail d'atelier et d'usine, est restée jusqu'à tout récemment purement empirique, sans aucun fondement dans les sciences cognitives (d'où le nouveau label ergonomie cognitive, pour désigner précisément une ergonomie adossée à l'étude des processus cognitifs). Comme jadis la physique a accouché les mathématiques, les technologies convergentes peuvent accoucher les sciences cognitives, leur faire produire des "applications" qu'elles seraient incapables, ou peu enclines, à imaginer. Elles leur offrent aussi des terrains inédits, situations nouvelles d'interaction, instruments de mesure, d'enregistrement et d'analyse incorporés aux outils nouveaux, modifications de systèmes biologiques naturels par prothèse ou hybridation (l'étude de la plasticité du cortex auditif après pose d'implant cochléaire donne une petite idée du genre de phénomènes entièrement nouveaux, sans équivalent "naturel", que produisent et produiront en nombre croissant les technologies convergentes).

D'autre part, les prophètes du "little bang" ne sont pas dans la même situation que les pionniers de la première IA : l'histoire bégaie peut-être, mais avec des modulations qu'on ne doit pas négliger. La première différence tient à la puissance des ordinateurs, et plus encore des réseaux, de demain; elle doit permettre notamment de compenser par imitation certaines faiblesses intellectuelles. Une deuxième différence réside dans l'"ubiquité" des dispositifs d'IA, rendue possible par l'ultraminiaturisation et le passage à la production de grande masse; cette ubiquité matérialise une idée qu'on a longtemps reproché à l'IA d'avoir ignorée, à savoir que l'intelligence est le résultat émergent de processus massivement parallèles (on dit, traditionnellement, que ces processus sont simples, mais en réalité les neurones n'ont rien de simple et la cognition réelle est peut-être l'effet d'un parallélisme massif de processus complexes, une réalité que nous n'avons pas encore les moyens d'appréhender conceptuellement; les « technologies convergentes » peuvent nous donner les moyens d'explorer empiriquement, à nos risques et périls, des situations de ce genre). Cette ubiquité débouche aussi nécessairement sur la constitution d'une seconde nature spécifique, dont internet et le téléphone portable nous fournissent un exemple simple mais frappant, seconde nature qui donne lieu à une dialectique d'appropriation et de confiscation aux profondes implications politiques mais aussi scientifiques.

Le stade ultime de cette dissémination des dispositifs d'intelligence artificielle est ce que désigne le label d'"intelligence ambiante" : lorsque ces dispositifs perdent tous les attributs de la machine, avec entrée, écran visualisant la sortie, unité centrale, et se fondent dans l'environnement, la soudure se fait entre le corps humain, seul ou en foule, d'un côté, le réseau de traitement de

l'information ambiante de l'autre, comme autrefois s'est opérée la soudure entre corps et vêtements, entre corps et outils, entre corps et canne d'aveugle, épée, baguette de chef ou plume d'oie; et entre foule et chemins, paysage, ville, réseau électrique... Mais cette fois, la "technosphère" ainsi constituée possède une réactivité, une cohésion et une autonomie impersonnelle entièrement nouvelles; elle est un organe social d'un genre singulier. Seule une démarche informée par les sciences cognitives peut permettre de sentir la texture de cette nouvelle toison de l'humanité et de la retailler dans le sens voulu.

(e) *Sciences cognitives appliquées et sciences de la cognition appliquée.* Que les sciences cognitives soient partie prenante de la convergence NBIC n'est que le signe le plus visible de leur applicabilité. Il existe désormais plusieurs domaines d'intérêt pratique auxquels les sciences cognitives sont en état de contribuer. Ceux qui sont généralement mis en avant sont l'éducation, l'éthique, la politique, le droit et l'économie, la recherche scientifique, l'art, les déficiences et autres pathologies mentales, motrices ou sensorielles, l'utilisation des artefacts en situation quotidienne, professionnelle, militaire. Ils ont d'ailleurs été, pour la plupart, déjà évoqués au passage, certains participant de près au "little bang".

Trois questions sont posées par cette expansion. La première est de savoir ce qui, dans la sphère des activités humaines, échapperait par principe à cette emprise ; on peut s'inquiéter à juste titre de la conclusion, irrésistible à première vue, du raisonnement qui part des deux prémisses que l'activité humaine résulte du déploiement des capacités cognitives de l'homme et que les sciences cognitives ont pour objet ces capacités. Raisonnement qui rappelle un peu celui qui fait des mathématiques la clé de la connaissance objective : tout ce qui est objectif se mesure ou se figure, or les mathématiques sont la science du nombre et de la figure, *ergo* les mathématiques sont la science fondamentale dont l'application aux différents domaines constitue l'ensemble des sciences, qui ne sont donc rien d'autre que des mathématiques appliquées. C'est du reste exactement une telle prétention universaliste, fondée sur la combinaison de ces deux inférences (la logique remplaçant, par restriction, les mathématiques), qui fascinait et repoussait dans l'IA de la première époque.

La deuxième question est comme l'envers de la première. En s'étendant, les sciences cognitives ne se diluent-elles pas en un cognitivisme faible, voire purement formel? Que les moutons se comptent ne fait ni de l'élevage ovin ni des troubles du sommeil une application des mathématiques.

En troisième lieu, on peut se demander où passe la frontière entre science pure et science appliquée, s'agissant de la cognition. Ici encore, on songe aux mathématiques. On connaît la boutade de ce mathématicien célèbre (André Weil) quand on l'interrogeait sur les rapports entre mathématiques pures et mathématiques appliquées : il n'y en a pas, répondait-il. Ce qui renvoie très exactement au problème soulevé : si une science fondamentale prétend sortir de ses frontières pour s'"appliquer" à une région de la réalité "concrète", rien ne peut l'arrêter, car toutes les régions de la réalité se valent, en ce qui la concerne. Et pour cette raison, elle cesse d'être elle-même. Elle ne devient science impériale qu'en renonçant à son être.

La solution que reçoit ce paradoxe dans le cas des mathématiques (dont on pourrait d'ailleurs considérer avec profit l'analogie pour la physique) est instructive. Les mathématiques appliquées existent bel et bien, ne se confondent pas avec la totalité des sciences, et entretiennent des rapports étroits avec les mathématiques pures. De fait, comme l'indique la réponse du grand mathématicien, elles *sont* des mathématiques, tout en se distinguant des parties pures du domaine par leur

orientation vers des structures concrètes : elles *visent* l'au-delà des mathématiques, tout en restant solidaires des mathématiques. De plus, il y a un gradient de pureté (ou d'application) entre les branches les plus abstraites et les branches les plus concrètes ; c'est à un continuum et non à une frontière précise qu'on a affaire. La genèse seule permet de faire une distinction nette, selon que le problème est posé de l'intérieur des mathématiques ou de l'extérieur. Mais telle branche appliquée peut donner naissance à des développements purs, qui sont alors une manière de mathématique appliquée pure, et le flou s'installe. Rien ne permet de caractériser à l'avance les domaines concrets que les mathématiques peuvent investir avec fruit. Enfin, le recours à des techniques mathématiques ne confère nullement à une science quelconque un lien de parenté avec les mathématiques : la biologie moléculaire n'est pas (dans son ensemble) une biologie mathématique, alors que la dynamique est une physique mathématique. La différence ne se laisse pas caractériser *a priori* : il faut aller voir de près quelles procédures mathématiques sont déployées, dans quel esprit, avec quelle profondeur conceptuelle et quelle productivité.

Revenons aux sciences cognitives. Il faut d'abord rappeler, même si c'est une évidence, qu'elles sont loin d'avoir atteint la maturité des mathématiques. Il n'empêche que les conditions de leur application ne semblent pas tellement différentes. En examinant un échantillon d'approches cognitives dans les domaines mentionnés plus haut, on observe un continuum qui va du cœur historique des sciences cognitives jusqu'à des thèmes qui se prêtent aussi naturellement aux sciences cognitives que l'arithmétique à l'élevage ovin, en passant par différentes nuances de cognitivisme de plus en plus dilué. Dans les meilleurs cas, on pioche dans une trousse à outils conceptuels pour proposer des distinctions nouvelles au sein d'une spécialité dont on respecte le cadre et l'architecture, et on en tire des éclairages inédits. Pour prendre un exemple, en droit ou dans d'autres disciplines normatives, il est intéressant de prendre en compte et d'étudier empiriquement les représentations que les agents (juges, justiciables) possèdent des concepts du domaine, l'apprentissage des règles et des normes, les raisonnements effectifs et non pas seulement reconstruits dans la phase de justification, etc. On ne saurait dire que des *résultats* soient transférés, ce qui est le propre d'une réelle application ; il s'agit, à l'heure actuelle, tout au plus d'heuristiques, ce qui n'est pas négligeable. Dans les cas les moins convaincants, on se contente de reprendre presque mot pour mot la présentation traditionnelle, en soulignant au passage les termes à contenu cognitif (tels que « mémoire » ou « raisonnement »), pour en conclure que le domaine est en passe de récolter les fruits du tournant cognitif. Ou alors, dans le pur style de la première IA, on propose une « formalisation » qui n'est fidèle ni sur le plan cognitif, ni sur celui du domaine considéré. Dans tous les cas, le tournant n'est pas affaire de décision, mais d'imagination scientifique et de travail empirique bien conduit.

Les sciences cognitives présentent un trait singulier, dont on ne voit pas d'équivalent en mathématiques (je laisse de côté la logique mathématique). C'est que la cognition elle-même, en tant que phénomène ou ensemble de processus, semble se prêter à la distinction pur/appliqué. La cognition pure serait constituée des processus généraux (mémoire, apprentissage, raisonnement, langage...) formant la « compétence » d'*Homo sapiens*, la cognition appliquée consistant à déployer et à développer cette compétence dans des types particuliers de situations réelles. Les partisans de la domanialité, ou modularité générale, rejetteraient sans doute la distinction, au motif que la « cognition pure » n'est qu'un autre nom de ce mythe qu'est l'« intelligence générale », objection qui me laisse sceptique : s'ils existent, les modules supérieurs sont en nombre bien plus faible que les types particuliers de situation, en sorte que le phénomène de l'application se retrouve nécessairement au niveau des compétences modulaires, sans changement notable. Une autre lecture de la distinction fait

de la cognition pure une variété de laboratoire, et de la cognition appliquée une variété sauvage : c'est le thème de la cognition située, évoqué plus haut, et dont le titre du livre de Hutchins (*Cognition in the Wild*) indique bien la teneur. Ce n'est pas le lieu d'entrer dans cette discussion, et je me contenterai de suggérer que les parties les plus solides des sciences cognitives appliquées sont, aujourd'hui, les sciences de la cognition appliquée : non pas projection sur un domaine de connaissance ou d'activité (architecture, droit, art, recherche scientifique...) de concepts, méthodes ou résultats des sciences cognitives pures, mais étude de processus proches de situations concrètes à l'aide de ces outils ; on en a vu des exemples plus haut.

(f) *L'interdisciplinarité*. Pour autant, faut-il séparer les bons des mauvais élèves, et ne juger qu'aux résultats ? Ce serait oublier qu'ils sont tous des élèves, plus ou moins ignorants, plus ou moins doués, mais tous inscrits à la bonne école, celle de l'interdisciplinarité. Les meilleurs ont su développer une pratique de l'interdisciplinarité qui combine la rigueur héritée des disciplines composantes et la fécondité d'interactions rendues possibles par la levée des barrières (comme cela s'est produit d'un bout à l'autre de l'histoire des sciences : rien de moderne dans cette démarche). Les moins bons ou ceux qui ont eu moins de chance dans le choix de leur sujet n'ont pas encore trouvé le passage, qui n'est jamais bien large. Mais ils introduisent déjà, comme on le voit par exemple dans les sciences de l'éducation ou en psychologie sociale, un regard, un tour de main, des distinctions, un style qui tranche. Pour revenir enfin à NBIC, la question est de savoir si la fameuse convergence saura trouver le chemin de l'interdisciplinarité bien comprise, ou alors s'égarera dans l'illusion d'une « interdiscipline » sans frontière, à la manière de l'IA du siècle dernier. Si la convergence n'est pas acquise, il y aura en revanche maints croisements et combinaisons dans le corps même des inventions techniques et aux interfaces avec les structures sociales, et de cette réalité-là il faudra bien faire et penser quelque chose.

II. Le nouveau visage des sciences cognitives

On peut distinguer, d'un côté, le répertoire des questions qui font l'actualité d'une science vivante, ainsi que les réponses partielles qui sont proposées et discutées, et de l'autre, les « matrices disciplinaires » dans lesquelles s'inscrivent les recherches. Cette distinction n'est ni nette ni robuste, mais elle est commode pour l'exposé synthétique des changements intervenus d'une période à la suivante. Dans cette seconde partie, ce sont les changements du second type qui sont examinés.

En regardant les choses de très loin, on est tenté de présenter le contraste entre la période 1960–1990 et la période dans laquelle nous sommes de la manière suivante.

Les sciences cognitives première manière reposent sur un équilibre entre cinq disciplines, IA, psychologie (cognitive), linguistique (générative et computationnelle), philosophie du langage et de l'esprit, neurosciences. Leur importance relative et leurs rapports mutuels sont régis par le cadre fonctionnaliste classique (séparation de principe entre niveau informationnel et niveau physique, effectivité des opérations du premier niveau, garantie par une stricte limitation des dispositifs physiques qui les réalisent ou en sont le siège au second niveau). L'IA se conçoit comme différente des autres : c'est une interdiscipline, qui constitue le liant conceptuel, méthodologique et technologique du champ, et c'est en même temps une science empirique dont l'objet est l'ordinateur, un artefact bien particulier puisqu'il est le modèle de l'esprit. Les quatre autres branches sont des fragments de disciplines classiques, nettement distingués du reste de ces disciplines : ainsi, la psychologie cognitive n'est qu'une petite province de la psychologie, de même la philosophie de l'esprit par rapport à la philosophie, la linguistique générative et computationnelle par rapport à la linguistique, les neurosciences cognitives par rapport aux neurosciences. L'IA s'attribue le rôle de vaisseau-amiral, rôle qu'on n'ose pas d'abord lui contester ouvertement ; la psychologie cognitive est en réalité, conceptuellement, au centre, mais elle professe un grand respect pour l'IA, dont les maîtres sont pour elle ses propres réformateurs, ou refondateurs ; la linguistique fraye sa voie sans beaucoup interagir, sur le terrain, avec les autres ; enfin, les neurosciences sont assez isolées, et ne pèsent pas très lourd, le fonctionnalisme strict ne leur laissant presque aucune initiative. Les philosophes, quant à eux (ceux, peu nombreux dans la corporation, qui prennent part à la « révolution cognitive »), volent, très raisonnablement, au secours du succès et parient dans l'ensemble sur l'IA, la grammaire générative et la psychologie computationnelle.

Cette image, caricaturale bien entendu, ne conserve sa validité relative qu'un assez bref moment, entre 1960 et 1975 environ, et la deuxième moitié de la première période est déjà celle de la transition vers une nouvelle configuration. La première édition du présent recueil (1992) pouvait donc présenter un bilan en forme de thèse et d'antithèse.

Loin d'être le moment d'une improbable synthèse, la seconde période est davantage celle d'une antithèse supérieure. Le fonctionnalisme strict, au sens historique, combinant, comme on y a longuement insisté dans l'introduction au volume et rappelé à l'instant, la séparation de niveaux information/réalisation physique, et la mise en équivalence d'opération mentale avec transformation calculable de données formelles, est mis de côté (je reviendrai plus loin sur le choix de cette locution). L'IA ne garde rien de son rôle de liant, que ce soit sur le plan conceptuel et méthodologique ou sur le plan technologique. Elle conserve bien, d'un côté, une branche théorique composite (faite de logique appliquée, de traitement de données linguistiques et plus largement sémantiques, de démonstration automatique et apprentissage symbolique, à quoi s'adjoint le territoire, annexé institutionnellement,

des méthodes formelles développées en réaction à l'IA classique, telles qu'algorithmes génétiques, réseaux, apprentissage non symbolique, ainsi parfois que vision algorithmique ou robotique), et de l'autre une branche technologique imbriquée dans l'informatique et, comme on l'a vu plus haut, dans les « technologies convergentes ».

La psychologie qui nous concerne ne renonce pas au qualificatif « cognitive », qui marque sa différence par rapport à la psychologie clinique d'inspiration psychanalytique, la psychologie différentielle, la psychologie sociale, la psychologie du travail et d'autres formes de psychologie appliquée. Mais elle s'étend à des domaines qui vont bien au-delà de son périmètre initial, et recoupe des thèmes qui sont également au centre d'autres branches de la psychologie —on en a vu quelques exemples. Enfin et surtout, sa structure se modifie profondément sous l'effet de deux facteurs. D'une part, la psychologie du développement et la neuropsychologie prennent une importance croissante ; elles ne sont plus des spécialités isolées, reposant exclusivement sur un vocabulaire, des méthodes et des habitudes professionnelles spécifiques, comme l'est encore par exemple, dans une grande mesure, la psychologie sociale ; elles s'approprient les concepts de la psychologie cognitive et la modifient en retour. D'autre part, la psychologie cognitive se rapproche des neurosciences au point où certains de ses représentants se considèrent désormais comme relevant de l'unique discipline des « neurosciences cognitives », au sein de laquelle les formations initiales (psychologie expérimentale ou biologie) comptent moins que les convergences sur un même objet.

Que cette fusion soit même concevable atteste un considérable déplacement du centre de gravité des sciences cognitives. Les neurosciences revendiquent désormais le rôle central, assumé notionnellement, dans la période précédente, par l'IA des débuts, puis par une psychologie computationnelle adossée au cadre computo-représentationnel. Le modèle de l'organe cognitif, c'est le cerveau lui-même, et c'est en étudiant et en modélisant l'organe, et non la fonction, que l'on élabore une science de la cognition. Cette formulation brutale, qui ne reflète rien des complexités de la situation réelle, suffit ici pour exposer la situation dans ses grandes lignes. Nous irons y voir de plus près dans un instant.

L'effet gravitationnel exercé par les neurosciences contribue à rapprocher les sciences cognitives de la sphère médicale. Le domaine d'intérêt commun est évidemment le cerveau et ses affections. La plupart des appareils d'imagerie cérébrale ont pour principale fonction d'aider au diagnostic neurologique, et sont implantés dans des services hospitaliers. Les psychologues cognitifs fréquentent ces services, et côtoient ainsi des médecins, et même des psychologues cliniques... Mais la médecine est impliquée à un autre niveau, celui des facteurs génétiques de maladies et syndromes neurologiques, et celui des troubles développementaux. S'introduisent de la sorte dans le voisinage des sciences cognitives la biologie cellulaire et moléculaire. Ici encore, les psychologues cognitifs sont amenés à fréquenter des gens qu'ils ignoraient, et qui en retour les méprisaient. Aux trois cultures présentes aux origines des sciences cognitives (sciences de la nature, sciences de l'ingénieur, sciences humaines) vient s'ajouter la culture médicale.

La biologie intervient également par une autre branche, la théorie de l'évolution, qui a considérablement progressé, elle aussi, au cours des vingt dernières années, tandis que les sciences cognitives se montrent beaucoup moins réticentes qu'au début à faire jouer les arguments évolutionnistes. On peut parler, à ce propos, d'une véritable coévolution des deux domaines.

La linguistique prospère, profitant de sa dynamique interne (extensions du domaine de la linguistique générative à de nombreuses langues, pluralité de modèles, développement de la sémantique et de la pragmatique) et des apports des sciences cognitives (logique, philosophie,

psycholinguistique, neuropsychologie, modèles neuromimétiques, traitement automatique, anthropologie, théorie de l'évolution,...). On peut même arguer qu'avec l'absorption graduelle de la psychologie dans les neurosciences, c'est la linguistique qui devient la discipline-carrefour, celle qui porte le projet initial des sciences cognitives.

Quant à la philosophie, elle est doublement stimulée par les difficultés rencontrées par la conception qu'elle avait élaborée pour servir de fondement ou de cadre unificateur des sciences cognitives : d'un côté elle étudie soit les moyens de sauver cette conception, soit les raisons de son erreur ; mais de l'autre, elle profite de l'autorisation que lui donnent les neurosciences de revenir à des thèmes (action, décision, émotion, corps, conscience...) et des traditions (divers courants de psychologie philosophique, phénoménologie, pragmatisme, philosophie des sciences...) qu'elle cultivait longtemps avant que n'émergent les sciences cognitives. Une bien plus grande variété de traditions et de ressources philosophiques trouvent désormais à s'investir dans le domaine, et touchent à la plupart des sujets qui le composent.

A. LES NEUROSCIENCES, VAISSEAU-AMIRAL?

(a) *L'irrésistible ascension de la neuroimagerie.* L'idée même de localiser certaines fonctions cérébrales par la mesure, et non l'inspection post mortem, remonte à près de quatre-vingts ans, lorsque l'électroencéphalographie (EEG) commença à être employée pour repérer les foyers épileptogènes et identifier les phases du sommeil. Le principe du tomodensitomètre (le « scanner ») est connu depuis quarante ans, et permet depuis plus de vingt ans aux neurologues de localiser lésions et tumeurs. Penfield, à Montréal, insérait dès les années 1930 des électrodes dans le cerveau de patients épileptiques pour stimuler des zones soupçonnées de receler les foyers épileptogènes.

Mais la neuroimagerie prend tout son essor au cours des années 1980, avec le développement du PET (ou *PET scan* : tomographe à émission de positrons), et puis surtout, une dizaine d'années plus tard, avec celui de l'imagerie fonctionnelle par résonance magnétique (ou IRMf, *fMRI* en anglais), de la résonance magnétique nucléaire fonctionnelle (RMN, *functional magnetic resonance spectroscopy* ou *MRS* en anglais) et d'une variante légère du PET, le *SPECT* (*single-photon-emission computerized tomography*). En parallèle, l'EEG, d'abord sans prise sur les processus cognitifs, se transforme en outil d'étude de la cognition (avec la méthode des potentiels évoqués, *ERP* ou *event-related potential* en anglais). ERP, et la magnéto-encéphalographie (ou MEG) viennent enrichir la trousse à outils de l'imageur. La stimulation magnétique transcrânienne (*TMS* en anglais), sans être à proprement parler une technique d'imagerie, puisqu'elle consiste à inhiber sélectivement et temporairement l'activité d'une aire du cerveau en la soumettant à l'action d'un champ magnétique traversant la boîte crânienne, se combine aux méthodes d'imagerie pour donner le même genre de résultat (la TMS est aussi une thérapie appliquée à titre expérimental en psychiatrie ; elle est non-invasive mais douloureuse). Enfin, l'enregistrement de l'activité de neurones individuels est en pleine expansion, en partie du fait de la possibilité, récemment apparue, d'enregistrer simultanément de nombreux neurones (une centaine, voire 240 dans certains montages : les techniques ne cessent d'évoluer, dans une atmosphère de compétition qui rappelle ce que connaissent la physique des particules ou la biologie moléculaire).

C'est peu dire que ces nouveaux outils ont modifié les pratiques et les perspectives des sciences cognitives : elles les ont bouleversées, autant peut-être que la lunette de Galilée et le

microscope de Loewenhoek ont fait de l'astronomie et de la biologie de leur époque. C'est du moins le sentiment d'une majorité de spécialistes des neurosciences cognitives, et de beaucoup de psychologues et de philosophes. Les données d'imagerie prennent une importance telle que l'observateur extérieur peut penser qu'elles sont l'alpha et l'oméga de la discipline. Une écrasante majorité de communications dans les congrès, et d'articles dans les revues, portent sur des résultats d'imagerie visant à établir l'« implication » de telle aire corticale dans l'exécution d'une tâche donnée, ou d'aires différentes dans l'exécution de deux tâches *a priori* semblables, ou encore l'intervention soit d'une modulation par un système donné, soit d'une activité psychologiquement significative, dans l'exécution d'une tâche. A simple titre d'illustration : le contrôle moteur implique-t-il un seul système prédisant le déroulement du mouvement et corrigeant les trajectoires, ou bien s'agit-il de systèmes, donc de processus différents ? L'imagerie semble plaider en faveur de la deuxième hypothèse. Autres exemples : l'attention sélective module-t-elle la réponse sensorielle aux stimuli, et par la médiation de quels mécanismes cérébraux ? Quels sont les processus cérébraux mis en jeu dans des tâches de catégorisation ? Les transformations syntaxiques postulées en grammaire générative possèdent-elles une réalité neurale ? Comment sont engendrés les sens nouveaux que prennent les mots dans des contextes inédits ? Quels sont les corrélats cérébraux de l'estime de soi ?

Dans le fascicule des contributions au congrès de la Cognitive Neuroscience Society, qui s'est tenu en avril 2004 à San Francisco, on trouve des résumés dépourvus de toute référence à l'imagerie, mais ils sont aussi rares que les rues sans pub dans le Dublin de Joyce (« Good puzzle would be cross Dublin without passing a pub. », *Ulysses*). Plus encore : c'est aussi le cas, à un moindre degré, des congrès de psychologie cognitive. Tout se passe comme si toute question portant sur la cognition devait être ramenée à une question de la forme : « Quelles aires du cerveau, ou quels composants du cortex, contribuent-ils au processus X, et comment leur action conjointe engendre-t-elle X ? », où X est un comportement ou un processus caractérisé dans le vocabulaire psychologique.

(b) *Critiques et objections.* Pourtant, comme Galilée et Loewenhoek, les « imageurs » contemporains se heurtent à des difficultés pratiques et théoriques et parfois au scepticisme.

Une partie de ce scepticisme provient d'une méconnaissance de ce dont les techniques d'imagerie, en très peu de temps, sont devenues capables. Si le PET souffre d'une résolution temporelle médiocre (90 secondes), l'IRMf est beaucoup plus précise, temporellement mais surtout spatialement. La résolution temporelle de la MEG est meilleure encore, et la combinaison, que l'on commence à mettre en œuvre malgré des obstacles considérables, permet d'obtenir dans certains cas une précision excellente dans les deux dimensions. Cela vaut aussi pour la combinaison entre potentiels évoqués (qui donnent la meilleure représentation possible de la succession d'événements neuraux au cours d'une tâche) et l'IRMf, qui peut remédier à l'indétermination spatiale de l'EEG (le « problème inverse » consiste à inférer la localisation de l'événement à partir de sa trace à la surface, et ce problème est « mal posé » : il n'admet de solution que moyennant des contraintes exogènes, que l'IRMf peut parfois fournir).

L'IRMf a d'autres vertus encore. Le cerveau entier peut être « photographié » en IRMf en une seconde, ce qui permet de se faire une idée de l'ensemble de ses activités au cours de l'exécution d'une tâche, car les traces hémodynamiques des réactions neurales subsistent pendant 4 à 8 secondes. Contrairement au PET, l'IRMf peut être pratiquée sans limite de durée ou de nombre d'examen, ce qui permet d'étudier des activités qui s'inscrivent dans une certaine durée, et des évolutions sur des semaines ou des mois.

Mais le PET comme l'IRMf ne mesurent l'activité neuronale que de manière indirecte, ce qui, s'ajoutant à l'extrême complexité des algorithmes de traitements de données, fait que les « photographies » des processus corticaux qu'elles procurent sont des reconstructions très indirectes. Le signal produit en IRMf en réponse à un stimulus est faible (de 0,1 à 1% de l'activité spontanée), et l'IRMf donne régulièrement des « faux négatifs », c'est-à-dire qu'elle conclut à la non-implication d'aires qui, selon d'autres techniques telles que la TMS, jouent en réalité un rôle. De plus, les variations interindividuelles sur le plan anatomique et sur le plan fonctionnel sont considérables, si bien qu'un moyennage sur 10 à 20 sujets, typique des expériences rapportées actuellement, n'est pas nécessairement significatif. Enfin, en l'absence d'une normalisation des repérages anatomiques, des mesures et des interprétations, la comparaison des innombrables expériences effectuées partout dans le monde est difficile, et conduit à une accumulation sans cumulativité.

Les enregistrements de neurones, quant à eux, qui semblent plus proches de l'activité proprement cognitive ou informationnelle du système nerveux, ne sont possibles que chez l'animal, et le transfert des résultats à l'homme pose des problèmes redoutables.

Ces difficultés suscitent des doutes qui s'expriment tantôt avec prudence, tantôt avec virulence, et qui s'adressent le plus souvent à une technologie particulière (venant des praticiens d'une technologie différente), ou parfois, au sein d'une technologie, à une école ou un laboratoire particulier, ou enfin à l'ensemble des techniques d'imagerie.

Arrêtons-nous un instant à ce dernier genre de critiques, qui semblent les plus menaçantes. Elles tournent toutes autour du même soupçon : l'imagerie ne produirait qu'une litanie de réponses sans questions, ou bien des réponses à des questions évidentes. Fodor, qui aime à se montrer, dans ce genre de polémique, plus obtus qu'il n'est, porte (par exemple dans la *London Review of Books* du 30 septembre 1999) les deux accusations : (i) les spécialistes des neurosciences semblent accorder du prix à n'importe quel résultat de la forme « Quand un sujet exécute telle tâche, telle partie de son cerveau est activée » (une réponse sans question scientifiquement plausible) ; (ii) tout résultat d'imagerie a la forme : « Quand un sujet exécute une tâche X, une partie du cerveau s'active qui est différente de celle qui s'active lorsqu'il exécute une tâche Y. » (une réponse à la question, évidente dès lors qu'on s'accorde à rejeter l'hypothèse de l'équipotentialité, en faveur d'une spécialisation des aires : « Deux tâches différentes mobilisent-elles des aires différentes du cerveau ? »). Bref, les neurosciences oscilleraient entre un empirisme forcené (mesurons, mesurons, nous en tirerons toujours quelque chose...) et une phrénologie élaborée sur le plan technologique, mais conceptuellement aussi fruste que celle de Gall et de Spurzheim (si je fais ceci, ça s'allume ici dans mon cerveau, si je fais cela, ça s'allume là). Pour conclure, Fodor indique clairement que ces recherches risquent de détourner des ressources matérielles et humaines de stratégies nettement plus prometteuses, celles d'une science cognitive bien comprise, c'est-à-dire pour l'essentiel, une psychologie cognitive computationnelle.

On peut facilement accorder au critique que bien des expériences d'imagerie sont insuffisamment préparées sur le plan théorique ; en privé, les bons « imageurs » se moquent de l'empirisme brouillon des mauvais. En imagerie comme dans n'importe quel paradigme expérimental, les bonnes expériences sont difficiles et par conséquent relativement rares, et il y a beaucoup de déchets. Mais cette concession à la sociologie scientifique ne compromet pas la défense sur le terrain théorique, une défense qui se déploie sur trois fronts.

Primo, Fodor n'a pas pris toute la mesure de la différence entre anatomie et physiologie, même si dans son étude classique de la notion de modularité, il distinguait le module comme aire

anatomique du module fonctionnel. Ce que l'imagerie révèle, c'est l'existence d'« actions à distance » : les aires forment des réseaux interactifs, et certaines fonctions cognitives sont le résultat, non de l'activité d'une aire particulière, mais de l'interaction de plusieurs. Les réseaux de neurones formels du connexionnisme postulaient une *distribution* de la représentation d'un concept ou d'un percept sur plusieurs unités, et dissociaient ainsi le niveau « manifeste » de la psychologie naïve du niveau scientifique de la ~~psychologie~~ ~~psychologie~~ (voir l'introduction et le chap. 2). Ce que les neurosciences, appuyées sur l'imagerie, laissent entrevoir, c'est une hypothèse plus hardie encore : les *fonctions* psychologiques seraient « distribuées » sur des ensembles de fonctions neurales, si bien que la psychologie naïve serait dissociée de l'ensemble articulé des fonctions « neurocognitives », ou, si l'on préfère, des traitements neuraux de l'information.

Mais n'est-ce pas ce que proposent depuis vingt-cinq ans les partisans du « matérialisme éliminatif », ces « neurophilosophes », Paul Churchland en tête ? Pas du tout : il ne s'agit pas d'éliminer les concepts de la psychologie ordinaire, surtout lorsqu'on n'a encore rien de solide pour les remplacer. L'objectif est d'établir progressivement des liens systématiques entre les deux niveaux de réalité, quitte à interposer des niveaux supplémentaires si nécessaire : pourquoi les neurosciences seraient-elles dispensées de mobiliser toute une série de niveaux, comme s'il leur fallait apporter une solution unique au problème des processus et au problème des deux aspects ? Ou encore, comme s'il suffisait de distinguer le niveau mental ou psychologique d'un niveau matériel fondamental (en l'occurrence celui du tissu neural) pour obtenir la description adéquate des processus physiques ? Quant à estimer possible dès à présent de proposer un langage de base, tel que celui des vecteurs d'activation que préconise Churchland, pour remplacer le vocabulaire intentionnel des représentations, cela relève d'une hubris philosophique au moins égale à celle du « réaliste intentionnel » qui jure que la science cognitive doit conserver les « croyances » et les « désirs » constitutifs, selon lui, de la psychologie ordinaire.

Ce que semblent proposer les neurosciences, en avançant l'hypothèse d'actions à distance et de multifonctionnalité des aires corticales, c'est de renoncer à la psychologie boxologique qu'acceptent, à presque deux siècles de distance, les phrénologues et les fonctionnalistes stricts, et de lui substituer une psychologie neurologique articulée d'un côté aux neurosciences fondamentales, de l'autre à la psychologie comme science du comportement au sens large, incluant les processus internes (mentaux). Il ne s'agit ni de conserver les notions de la psychologie philosophico-naïve à travers des niveaux d'implémentation (une notion qui, soit dit en passant, interdit à peu près toute authentique découverte en neuroscience cognitive ou fonctionnelle), ni d'éliminer par principe ces notions au profit de catégories 'neurocomputationnelles' essentiellement spéculatives qui anticipent sur la découverte. Churchland et Fodor sont renvoyés dos à dos.

Secundo, les effets de bonne fortune (*serendipity*) qui sont, pour Fodor, la marque d'une science sans principes, ont tout au contraire un rôle essentiel à jouer, comme toujours dans les sciences et tout particulièrement lorsque, comme les sciences de la vie, elles ont pour objet des systèmes complexes. En l'occurrence, supposons que nous remarquons que l'exécution de la tâche A met en branle une aire dont nous avons précédemment acquis la conviction qu'elle est impliquée dans l'exécution de la tâche B (en un sens très étendu : un processus quelconque, qui n'est nécessairement ni conscient, ni moteur), alors que rien ne nous disposait à penser que A et B ont le moindre rapport entre eux. Cela ne nous ouvrirait-il pas des horizons nouveaux en psychologie même ? Ou supposons inversement que deux tâches, que nous avons l'habitude de considérer comme étroitement apparentées, tout en les distinguant, aient des signatures très différentes en imagerie : n'allons-nous

pas mettre en question l'évidence de leur proximité psychologique, quitte à nous interroger aussi sur les raisons de cette apparence ?

Or nous avons des exemples de découvertes de ce genre, il y en a très probablement déjà beaucoup, et on peut parier qu'elles vont se multiplier. Le cervelet, dont on a longtemps pensé que le rôle se bornait à la coordination motrice, est « impliqué », au sens ordinaire comme au sens technique, dans une série de processus cognitifs touchant à l'attention, à l'imagerie mentale, au contrôle de la succession temporelle etc. Que ces fonctions « partagent » avec le contrôle moteur une ressource aussi massive que le cervelet ne peut pas ne pas être pris en compte. Dans les premières tentatives pour éclaircir cette question, on a proposé une distinction entre les tâches exigeant un contrôle précis de la succession, et celles qui peuvent être déclenchées dans des fenêtres temporelles larges. Une telle distinction ne s'impose pas d'emblée dans l'analyse conceptuelle ; quant à l'idée que la succession précise fasse appel à un organe particulier, qui n'intervient pas pour réguler des processus à succession grossière, c'est sûrement une découverte aussi peu accessible à l'analyste que le fait que la préhension soit assurée chez l'éléphant par la trompe, ou la miction chez le mammifère mâle par l'organe reproducteur : le bricolage échappe aux gros doigts de l'analyse conceptuelle.

Un autre exemple est fourni par l'étude en imagerie des processus de catégorisation. Les psychologues débattent depuis deux décennies des mérites comparés de trois théories de la catégorisation. Pour la première, catégoriser un nouvel objet, c'est-à-dire répondre à la question de savoir s'il appartient à une catégorie donnée, c'est lui appliquer des règles ; pour la deuxième, c'est le comparer à un « prototype » de la catégorie ; pour la troisième, c'est le comparer à certains représentants particuliers de la catégorie conservés en mémoire. En croisant une série de résultats d'imagerie, on donne corps à l'hypothèse que chacune des théories décrit un processus réel, réalisé dans le cerveau de manière différente.

L'exemple le plus spectaculaire sans doute est celui des neurones-miroirs, dont il a été question plus haut. Il semble bien qu'ils aient été découverts par hasard (au sens particulier que prend le hasard en matière de recherche scientifique : un ignorant, si chanceux soit-il, ne peut évidemment rien découvrir en science—je laisse de côté cette question ressassée, qui n'affecte pas spécifiquement l'imagerie cérébrale fonctionnelle). L'équipe de Giacomo Rizzolatti, à Parme, enregistrait l'activité de neurones individuels dans l'aire F5 du cortex d'un macaque, connue pour être impliquée dans les actions de la main et de la bouche ; un des neurones sous surveillance se mit un jour à crépiter sans que le macaque eût mû sa bouche ou sa main. Le facteur déclenchant était le *spectacle* d'un expérimentateur portant avec sa *propre main* de la nourriture à sa *propre* bouche. Il aurait fallu une imagination considérable au psychologue-philosophe pour conjecturer une pure et simple identité entre le processus de reconnaissance, chez autrui, d'une séquence motrice réalisant une intention donnée, et l'exécution d'une semblable séquence par le sujet.

Ce n'est pas dire que ce soit absolument impossible, ni d'ailleurs que l'hypothèse soit d'emblée pleinement intelligible et au-dessus de toute critique, tant s'en faut. De manière générale, les résultats obtenus par neuroimagerie prêtent le flanc à toutes sortes de critiques. Mais la vertu essentielle de la neuroimagerie demeure : elle ruine la vision dogmatique du psychologue-philosophe, et met un terme à ce qu'on pourrait appeler le « monopole turingien ». Rappelons-nous la prouesse de Turing, qui par la seule analyse conceptuelle met au jour une base de ressources fondamentales pour l'exécution d'un certain type d'opération mentale. L'illusion de Fodor, c'est que cette méthode, renforcée par le recours à la psychologie expérimentale, suffit à identifier le répertoire complet des ressources élémentaires suffisantes pour engendrer tout processus cognitif. Or rien ne permet de le

penser. Mais même à supposer (contre toute vraisemblance) que l'analyse turingienne (au sens étendu que je propose) ait théoriquement la puissance requise, s'interdire le recours à d'autres méthodes est pragmatiquement absurde : si l'on peut sans doute traverser le Sahara à cloche-pied, ce n'est pas conseillé à qui veut arriver en un temps raisonnable.

La neuroimagerie est-elle la lunette de Galilée des sciences de l'esprit-cerveau ? Si c'est ce cas, refuser d'y recourir, en raison des difficultés méthodologiques et conceptuelles qui s'attachent à son usage, c'est faire comme certains contemporains de Galilée. Une attitude qui n'est nullement dénuée de justification, mais qui cesse rapidement d'être tenable. Si l'on veut contester le parallèle, il faut, je pense, rejeter l'idée même d'esprit-cerveau, et épouser une forme de dualisme. Mais il semble difficile de maintenir un naturalisme intégral et en même temps d'imposer des limites ou une forme a priori à la contribution des neurosciences.

(c) *Au-delà de l'imagerie.* Si la lunette de Galilée a finalement eu raison des résistances, c'est aussi parce que son inventeur a eu la sagesse de n'en faire ni l'arbitre suprême, ni même le fournisseur exclusif de données empiriques. Les spécialistes les plus avisés des neurosciences ne voient pas dans l'imagerie autre chose qu'un outil, certes puissant, parmi d'autres et ils savent mieux que quiconque en dehors de leur profession que, comme Duhem fut l'un des premiers à souligner, les instruments des sciences avancées sont eux-mêmes des théories incarnées ; et que les résultats des protocoles d'imagerie sont le fruit d'amalgames complexes d'observations et de conjectures, lesquelles confrontent le tribunal de l'expérience solidairement avec les théories élaborées à partir des mesures livrées par les instruments.

Les neurosciences débordent largement l'imagerie, c'est une évidence. L'implication de différents ensembles d'aires dans différentes tâches peut même apparaître comme un aspect mineur de la science du cerveau. La place qui est faite néanmoins actuellement à l'imagerie s'explique en partie, paradoxalement, par le fait qu'une partie considérable de la discipline, qui ne doit rien à l'imagerie, est un acquis. L'effet de nouveauté joue à plein : les recherches menées dans les domaines établis font moins parler d'elles, tandis que les chercheurs veulent « essayer » les nouvelles techniques de l'imagerie. Pour beaucoup de spécialistes du cerveau, comme pour les journalistes et les philosophes, la connexion apparemment directe avec la psychologie donne à l'imagerie un attrait irrésistible. Mais cet enthousiasme n'est pas partagé par tous. L'essentiel est que l'imagerie n'est *pas* le *pilote* du vaisseau-amiral, qu'elle n'est qu'un outil ; et que cet outil n'est pas au service exclusif de ce qu'on pourrait appeler, d'un terme légèrement trompeur, la « neuropsychologie fonctionnelle ». Elle est mise à profit, par exemple, par deux branches essentielles des neurosciences, l'anatomie et l'étude du développement. Des questions aussi décisives que celles du codage neural ou du rôle et du fonctionnement des différentes catégories de cellules (non seulement les neurones, mais également les cellules gliales), sont moins directement affectées par les nouvelles techniques d'imagerie. Mais surtout, ce que les neurosciences recherchent, ce sont des hypothèses structurales, au sens théorique (et non anatomique) du terme, et les disciplines partenaires ont tout autant que les spécialistes des différentes neurosciences vocation à formuler de telles hypothèses (qu'on songe en particulier à la dimension évolutionnaire). Dans cette perspective, les « régions » que sont les programmes de recherche jouent le rôle moteur, plus que les « nations » que sont les disciplines institutionnelles, au sein d'une « Europe » des sciences cognitives.

Ne sommes-nous pas revenus à notre point de départ, c'est-à-dire à la configuration fondatrice des sciences cognitives ? Pas vraiment, puisque, encore une fois, le cadre des relations entre

les disciplines a véritablement éclaté. Prenons une fois encore l'exemple des neurones-miroirs. Les neurosciences fournissent, dans un premier temps, un couple fonction/aire chez l'animal : mouvements de la main/aire F5 du cortex prémoteur, et une théorie d'arrière-plan (à ne pas oublier) sur la chaîne causale qui relie un neurone prémoteur aux mouvements de la main et de la bouche ; elles exploitent un cadre de psychologie animale pour définir et contrôler un comportement particulier. La chance qui ne sourit qu'aux bons et patients chercheurs (Rizzolatti travaillait depuis 15 ans sur la question du contrôle de la main du macaque) conduit à l'hypothèse d'un système « mimétique » intervenant tant dans l'observation que l'exécution d'une action (au sens intentionnel du terme). Ce n'est qu'à ce stade que les philosophes et les psychologues interviennent, non du tout pour imposer une grille d'analyse computationnelle, mais pour soumettre l'interprétation proposée à une étude critique et pour formuler des questions susceptibles de la mettre à l'épreuve. La question se pose du transfert à l'homme, et ce sont à nouveau les neurosciences qui mènent le jeu, fortes de leur savoir acquis sur l'homologie entre l'aire F5 du macaque et l'aire de Broca chez l'homme. A nouveau psychologues et philosophes contribuent à l'enquête, grâce cette fois à l'expérience acquise dans la compréhension de concepts tels que représentation, action, intention, simulation, ... notamment au cours des travaux menés essentiellement par ces psychologues et philosophes sur la « théorie de l'esprit » (voir I.B.a *supra*). Rien dans ces allers-retours ne repose sur le cadre fonctionnaliste strict, ni n'étaie, du moins directement, l'idée que l'étude du cerveau est à celle de l'esprit ce que l'électronique est à l'informatique. Rien n'encourage l'espoir, entretenu par l'IA de la première époque et par les philosophes intéressés alors, que les fonctions postulées par l'équipe de Parme résultent de la composition de « primitives » figurant dans un répertoire qu'ils ont, ou auraient pu, établir par leurs propres moyens. (On peut en dire autant des lois d'évolution déterminant les trajectoires des états des systèmes et sous-systèmes cérébraux, mais je laisse de côté cet aspect, pourtant essentiel, car il engage des difficultés techniques.)

Ce n'est pas la première fois que ce genre de situation se produit. Laissant de côté les cas (semblables dans leur structure) où les neurosciences étaient déjà les fournisseurs de fonctions non prévues, voire non prévisibles, par l'analyse fonctionnaliste, on peut penser à la linguistique, à la psychophysique et aux théories computationnelles de la vision, aux modèles connexionnistes. Les fonctions postulées par ces disciplines pour rendre compte de phénomènes liés à des facultés aussi centrales que le langage, la vision, la mémoire, la catégorisation, étaient sans équivalent dans le vocabulaire de la psychologie philosophico-naïve, d'où la proposition de les regrouper sous l'étiquette générale de « subdoxastique » ou « subpersonnel ». Oui, mais... : mais ces disciplines n'ont pas l'ancrage dans le « réel » dont les neurosciences se prévalent. Elles sont abstraites, conjecturales. Les neurosciences décrivent un objet physique et biologique bien déterminé, au foyer convergent de la théorie de l'évolution, de la génétique, de la neuroanatomie, de la biologie cellulaire, etc. C'est en tout cas comme cela que les neurosciences ont tendance à voir les choses.

(d) Résultats et promesses : les leçons de l'IA. Et c'est exactement à ce point que se fait sentir la tentation de considérer le fonctionnalisme sous toutes ses formes comme réfuté, et de vendre l'ensemble de l'entreprise des sciences cognitives à sa puissante filiale des neurosciences, appuyée sur la maison-mère de la biologie. A cette tentation, il faut, je crois, résister.

Il ne s'agit ni d'un problème purement institutionnel, ni d'une question purement verbale. Certains diraient sans doute en effet que c'est la « réalité » des recherches qui compte, que l'étiquette disciplinaire est sans importance ; ou bien alors, qu'elle en a une, mais d'ordre seulement

sociologique, avec des conséquences concernant le recrutement des étudiants, le financement, les critères d'évaluation, etc. On peut penser au contraire que la question de savoir de qui le cerveau est l'affaire est essentiellement de nature scientifique, et que la réponse « Des biologistes, naturellement » est courte. Il suffit d'ailleurs de la poser avec cette brutalité pour le voir. Qui dirait que le corps est l'affaire des seuls médecins, ou la société celle des seuls sociologues ?

Pour le dire rapidement, et pour éviter de reprendre à nouveaux frais cette question, la question, posée dès avant l'apparition des sciences cognitives, notamment par les premiers cybernéticiens, la raison de ne pas laisser le cerveau entre les mains des seuls biologistes tient en un mot : complexité. Les biologistes n'ont un privilège de compétence exclusive que dans un certain mode d'analyse, propre à élucider certains aspects, à produire certaines sortes de description qui, en raison de la complexité de l'objet, ne peuvent pas épuiser ce que nous voulons en savoir. Mais, demandera-t-on (en feignant la naïveté), quels sont donc ces aspects ou niveaux qui échapperaient par définition à l'œil du biologiste ? La complexité à elle seule oblige-t-elle à recourir à plusieurs disciplines ? La difficulté est-elle, à l'ordre de grandeur près, celle que pose, mettons, la main, dont la compréhension exigerait les lumières non seulement de nombreuses branches des sciences bio-médicales, mais aussi celles de l'anthropologie, de l'archéologie, de l'histoire de l'art... Ou est-ce la complexité purement *physique* qui fait du cerveau un organe différent, du point de vue qui nous concerne ici, du foie ou du rein ? Ou bien faut-il renoncer à toute comparaison, admettre qu'il n'existe rien dans l'univers qui soit à la fois aussi important et aussi compliqué que le cerveau, et constater que, jusqu'à présent, les biologistes ne nous ont pas convaincus qu'ils étaient de taille à l'affronter seuls ? Ou encore, invoquer l'idée banale de niveau de réalité, ou celle, proche, de structure érotérique de la science : selon la question que l'on pose au sujet du cerveau, différentes sciences auraient vocation à fournir la réponse. Le linguiste demande comment le cerveau produit la capacité de parler, le psychiatre demande quel médicament peut soulager les symptômes de la schizophrénie, le sociologue demande si l'intelligence est « génétiquement déterminée », le kinésithérapeute demande si l'on peut réparer un cerveau lésé pour permettre à son patient de marcher, le pédagogue veut savoir s'il faut enseigner deux langues vivantes dès l'âge de trois ans, le physiologiste cherche à comprendre comment l'énergie est fournie au cerveau, l'anthropophage demande combien de temps il faut cuire un cerveau au courbouillon, etc.

Je n'essaierai pas de poursuivre ces différentes pistes. Je préfère me tourner vers l'ordinateur, qui est souvent présenté comme incarnant la bonne réponse : ce système physique parfaitement concret est en un sens parfaitement connu (notamment par son constructeur). Laisant de côté le fait qu'en réalité il pose encore à la science des questions difficiles, demandons-nous comment un groupe de savants venus de Jupiter pourraient s'y prendre pour en saisir le fonctionnement. Confieraient-ils la tâche à leurs seuls physiciens ? La réponse attendue, comme au théâtre des marionnettes, est évidemment « Non ! » Sans une idée relativement précise des fonctions remplies par l'objet, la détermination de sa structure serait une entreprise désespérée. Réponse que le neurobiologiste rejettera peut-être, au motif qu'elle repose sur une comparaison entre un système naturel et un objet artificiel, façonné par l'esprit humain à sa propre image. Il lui sera difficile néanmoins de contester deux choses : (i) que même dans le cas de systèmes infiniment plus simples tels que le cœur ou le pancréas, le biologiste doit mener de front l'élucidation de la structure et celle de la fonction ; (ii) que contrairement à ces cas simples, où les deux chantiers sont menés par différents spécialistes au sein de la biologie, « la » fonction du cerveau ne semble pas être du ressort exclusif des biologistes, à

moins que ceux-ci n'incorporent la totalité des spécialités concernées par la cognition, ou encore qu'ils se muent eux-mêmes en linguistes, en logiciens, en psychologues, etc.

Cette discussion, même à l'état d'ébauche sommaire, provoque un fort sentiment de déjà-vu. Elle reproduit point par point le débat critique autour de l'IA de la première époque. Ses défenseurs prétendaient qu'elle avait vocation à élaborer une théorie générale et exhaustive de l'intelligence *naturelle* (la simulation n'étant qu'un moyen d'investigation, et une application). On leur demandait comment ils estimaient pouvoir se passer des spécialistes de la structure responsable de la fonction dont ils faisaient la théorie ; leur réponse, on le sait, était le fonctionnalisme. On leur demandait ensuite d'où ils tenaient leur connaissance précise de la fonction ; ils répondaient qu'ils étaient également psychologues, logiciens et linguistes...

L'échec de leur stratégie maximaliste devrait nous renseigner sur les risques d'un « biologisme » exclusif concernant le cerveau. D'abord, paradoxalement, l'affaiblissement du fonctionnalisme prive de son efficacité la parade du biologiste « impérial » : s'il prétend que la cognition se négocie, mais que le cerveau lui appartient, il s'expose à l'objection que cette séparation n'est plus possible, justement parce qu'elle reposait sur une conception discréditée des rapports entre structure et fonction (la neurophilosophie, en ce sens primitif, est donc auto-réfutante). Ensuite, si nous comparons la situation dialectique actuelle à celle des années 1970, nous remarquons une symétrie presque parfaite entre la position de l'IA alors et celle des neurosciences aujourd'hui. *Dans toute cette discussion, il n'est question que des versions « intégristes » des doctrines* : il est parfaitement évident que la plupart des pratiquants de l'IA d'hier, et des neurosciences d'aujourd'hui, ne défendent pas les positions que je réunis ici dans un même regard critique.

L'IA maximaliste d'hier, comme les neurosciences impériales d'aujourd'hui, estiment qu'il leur revient de fixer le sujet d'étude principal : l'intelligence (comprise comme l'ensemble des fonctions cognitives) pour la première, le cerveau (le système nerveux central) pour la seconde. L'argument est le même : ce qui est réel, concret, sur quoi l'enquête scientifique a prise, c'est l'intelligence, respectivement le cerveau. Dans les deux cas, une large place est faite à l'interdisciplinarité : toutes les bonnes volontés sont les bienvenues, car la tâche est immense ; mais dans les deux cas, le maître d'œuvre est clairement déterminé : c'est la discipline qui sait ce dont il s'agit *au fond*. Dans les deux cas, cette conviction implique un critère de scientificité : une bonne théorie, pour l'IA, doit conduire à un (ou plusieurs) *programmes* (au sens informatique du terme) ; pour les neurosciences, à la description d'un processus cérébral, d'un codage ou d'un circuit neuronal, en principe visualisable par imagerie. Dans les deux cas cependant, les « solutions » offertes posent souvent autant de questions qu'elles ne fournissent de réponses. Dans les deux cas, une désapprobation, voire davantage, s'attache à tout projet de recherche qui n'inclut pas un volet relevant de la discipline maîtresse : il n'appartient pas au domaine, il manque de sérieux scientifique, il relève d'un stade dépassé des connaissances, il est symptomatique d'un risque de régression. Dans les deux cas, les premiers succès sont célébrés sans retenue, et vus comme annonciateurs d'une victoire finale à terme relativement rapproché. Dans les deux cas, une attitude critique est considérée comme malvenue sur le plan de la politique scientifique, même si elle est intellectuellement acceptable, car la discipline maîtresse ne l'est pas chez elle : la position de l'IA au sein de l'informatique était minoritaire et fragile, comme est celle des neurosciences, surtout intégratives, au sein de la biologie dominée par la biologie moléculaire, la génétique et la (post) génomique ; dans ces conditions, une critique de la part des disciplines alliées est quasiment une trahison.

J'admets sans difficulté que la symétrie a ses limites. Il me paraît probable que les neurosciences sont plus solidement ancrées, et ont davantage d'avenir, que l'IA prométhéenne des années 1960–70. En tout état de cause, l'IA n'a pas bénéficié d'une période suffisamment longue de prospérité théorique pour accumuler un savoir comparable en richesse à celui des neurosciences. On ne pourra me reprocher d'avoir, dans les pages qui précèdent, minimisé la contribution des neurosciences, ni d'avoir jadis épargné mes critiques à l'IA. Il faut pourtant dire sans détour que les ambitions de l'IA n'étaient à l'époque pas dénuées de justification, que la conception « symbolique » de la cognition qu'elle mettait en œuvre à sa façon conserve en psychologie une pertinence éminente ; qu'inversement les neurosciences sont loin d'être en mesure aujourd'hui d'éclairer de façon décisive toutes les grandes questions relatives à la cognition, que dans de nombreux cas leur contribution est fortement tributaire des autres disciplines, qu'elles ne sont pas, ou pas encore, en mesure ne fût-ce que d'aborder certains problèmes, et enfin qu'elles sont en attente d'hypothèses structurantes. Cela ne diminue pas d'un iota leur mérite et la beauté des résultats qu'elles ont obtenus, et dont le philosophe que je suis n'a certainement pas pu prendre toute la mesure. Ces considérations ne visent qu'à une chose, mettre en doute l'idée que les neurosciences sont à elles seules dès à présent le cœur des sciences cognitives. Le domaine reste à mon sens « distribué » sur une grande variété de disciplines et de sous-disciplines, et conserve la structure en anneau qu'il a depuis sa naissance, un anneau dont le centre géométrique n'est occupé en permanence par personne.

B. OÙ EST PASSÉ LE MAINSTREAM?

(a) *Les difficultés philosophiques du fonctionnalisme.* On a évoqué à plusieurs reprises la « mise à l'écart », l'« affaiblissement » du fonctionnalisme. On n'a pas parlé de déroute ou de rejet complet. L'esprit de prudence aurait-il une fois encore frappé ?

Non : ces expressions reflètent une ambiguïté, et une incertitude, qui sont inhérentes à la situation. L'ambiguïté est celle-ci : le fonctionnalisme en est venu à désigner deux doctrines distinctes, quoique souvent associées. La première est essentiellement méthodologique, et reprend l'idée générale de fonctionnalisme qui court à travers les sciences et les époques, tout en lui donnant une tournure caractéristique. En ce sens, le fonctionnalisme consiste à préconiser une approche des phénomènes par la fonction, et non par les mécanismes matériels en vertu desquels la fonction est réalisée (ce qu'on appelle traditionnellement en biologie la « structure »). La tournure caractéristique vient de la nature que l'on attribue à la fonction : dans les sciences cognitives, cette fonction est *informationnelle*. La cognition est, selon cette doctrine, affaire de « traitement » de l'information. Un organe est cognitif (plutôt que digestif, respiratoire, métallurgique, politique, culturel) dans la mesure où il tire sa raison d'être du fait qu'il reçoit, stocke et renvoie de l'information ; c'est-à-dire qu'il est invariant par un certain ensemble de transformations physiques qui préservent l'information (une quantité abstraite, comme peut l'être l'entropie en thermodynamique) : l'organe se comporte de la même façon si l'on remplace tel transfert d'énergie par un transfert équivalent sur le plan informationnel.

Mais qu'est l'information ? De quelles transformations parlons-nous ? C'est à répondre à cette question de manière précise, ou du moins d'une façon qui permette de mettre en route un programme de recherche, que sert la deuxième doctrine fonctionnaliste. Ce fonctionnalisme « strict », comme on l'a appelé plus haut, est le computationnalisme turingien, mieux nommé *computo-*

représentationnalisme, abondamment exposé dans l'introduction au présent volume et de manière plus complète dans le chapitre 12. L'information est le contenu sémantique ou intentionnel commun à divers « véhicules » matériels, de la manière dont le son « a » est le contenu phonologique commun à toutes sortes d'événements sonores, de graphies, voire de souvenirs (mon souvenir du « Ah ! je ris de me voir si belle en ce miroir »), ou de la manière dont le nombre 4 est attaché à toutes sortes d'ensembles (les chevaliers de l'Apocalypse, les points cardinaux,...), de graphies (« 4 », « IV »,...), de pensées, etc. Les processus causals affectant certaines catégories d'entités ne sont sensibles qu'à leurs contenus informationnels, en sorte que le quotient de ces processus par la relation d'équivalence opère sur ces contenus ; enfin, ces opérations quotientées sont effectives, au sens de la computation turingienne, dont on sait par ailleurs la stabilité définitionnelle. Tout cela forme un tout théorique cohérent, symbolisé par l'ordinateur, ou encore par le « pont de Londres » dont la travée inférieure est une « machine syntaxique » (un système physique de transformation de stimuli) mimant la travée supérieure, une « machine sémantique », c'est-à-dire un système abstrait de traitement de l'information, les piliers étant fournis par la sémantique des entités physiques, et la stabilité par l'invariance sous toute transformation équi-informationnelle.

Ce cadre théorique à deux strates est l'objet de plusieurs formes de remise en cause. La première concerne la cohérence du fonctionnalisme strict avec une conception générale de l'explication scientifique. La deuxième vise la pertinence du fonctionnalisme strict comme mise en œuvre du fonctionnalisme général. La troisième porte sur la fécondité du fonctionnalisme général comme cadre pour les recherches sur le cerveau et/ou l'esprit.

Le premier souci est d'ordre philosophique, et il va en être question dans un instant. Les deux autres soucis sont d'ordre empirique : ils concernent, pour le premier, la validité du modèle turingien (dans le cadre d'une approche informationnelle de la cognition), et pour le second, l'opérativité du concept d'information, ou encore l'utilité, pour les sciences cognitives, de recourir au vocabulaire informationnel. Dire que le fonctionnalisme est rejeté peut donc signifier, soit qu'il a été jugé contradictoire ou confus sur le plan conceptuel (ce qui ne peut s'appliquer qu'à sa version stricte), soit qu'on s'est accordé sur son inadéquation matérielle à son objet, le cerveau/esprit (selon les cas, sera visé le fonctionnalisme turingien ou bien tout fonctionnalisme apparenté), soit enfin que son utilité méthodologique est jugée nulle. En parlant plus haut de « mise à l'écart », je visais à la fois la deuxième et la troisième éventualité : le progrès des neurosciences d'une part, le développement du « secteur libre » des sciences cognitives (les recherches apparemment indifférentes à la question du fonctionnalisme), d'autre part, ont en pratique détourné l'attention de la question de la validité d'un modèle computationnel de la cognition, quel qu'il soit (classique, connexionniste de telle ou telle variété), et partant, ont affaibli la position du fonctionnalisme général comme « programme métaphysique de recherche » ou « hypothèse structurale » pour les sciences cognitives.

Ayant ainsi défendu ma caractérisation nuancée de la situation générale du fonctionnalisme, je dirai quelques mots de ses difficultés internes. Je mentionnerai les deux principales. La première est de concilier le fonctionnalisme strict avec le développement présent des neurosciences cognitives, qui semblent ne rien lui devoir, ou même lui opposer un contre-exemple. A ce problème qui est de l'ordre du constat se rattachent des mises en cause directes du computationnalisme : le constructivisme, le dynamicisme se posent en doctrines rivales. La seconde difficulté concerne la « préemption causale » ou épiphénoménalisme que Pierre Jacob évoque dans le chapitre 12, et que le philosophe Jaegwon Kim n'a cessé d'affûter au cours des dernières années. Selon Kim et d'autres, la clôture causale de la physique est incompatible avec la causalité mentale ou intentionnelle. Ainsi, l'explication par les

causes physique de mon action de lever le bras prive de pertinence l'explication par la cause mentale (mon intention de lever le bras, mon désir d'attirer l'attention du président de séance). La réponse qui vient à l'esprit est que ce désir est matérialisé par l'une parmi un ensemble de configurations physiques différentes, et qu'il se réalise (en tant que désir) par l'effectuation des causalités matérielles relatives à cette configuration particulière. Mais un schéma causal n'est pas plus une cause que l'est par exemple une équation d'évolution : c'est une explication sans effet sur le monde. Or l'idée de cause mentale appelle bien semble-t-il une notion littérale de cause : en signalant ma présence au président de séance, je ne me contente pas de renvoyer platoniquement à une cause physique particulière (déclenchant une cascade d'activations neuronales), je cause, ou plutôt mon désir cause, le mouvement de mon bras. La difficulté est là. Elle doit être levée, faute de quoi un réductionnisme intégral (défendu par Kim et par Bickle) semble menacer, ce qui n'est sans doute pas propice aux sciences cognitives (qui ne sont après tout que l'une des « sciences spéciales », comme la géologie ou l'embryologie, et ne devraient pas bénéficier d'un statut inférieur).

Ce n'est pas le lieu de poursuivre cette discussion. Je partage pour ma part avec certains philosophes des doutes sur son bien-fondé ; les notions de causalité et d'explication causale sont probablement à l'origine de confusions. Ce qu'il faut retenir, c'est le fait que les philosophes de l'esprit eux-mêmes s'inquiètent de cette objection, et d'autres encore, concernant le plus souvent la perception, l'action, le corps, les émotions, la conscience, bref, des phénomènes, éloignés en apparence du calcul ou du raisonnement, pour lesquels la distinction entre information et véhicule de l'information semble problématique.

(b) *La cause et le calcul.* Quoi qu'il en soit de ces difficultés philosophiques, elles ont pour effet et aussi peut-être en partie pour cause un désinvestissement, de la part des scientifiques eux-mêmes, de la notion de calcul au profit de la notion de cause. En d'autres termes, le gain promis par le fonctionnalisme, à savoir la mise entre parenthèses des causes particulières, les sciences de la cognition semblent bien moins que dans une phase antérieure en mesure d'en bénéficier. Les questions qui mobilisent leur attention concernent bien davantage les causes et les mécanismes que la caractérisation abstraite des processus comme calculs. Ce qui est conservé de la période précédente, c'est le vocabulaire informationnel : on continue de parler de l'information qui circule entre telle aire et telle autre, plutôt que de dépendance causale, et la raison en est, tout simplement, qu'il est impossible, ici comme ailleurs en biologie, de se passer de téléologie : une cause nue est sans fonction, une transaction informationnelle remplit une fonction pour l'organisme qui en est le siège. En ce sens, l'orientation biologique du domaine est incontestable, et dépasse la stricte contribution des neurosciences. La plupart des travaux en cours se présentent comme autant d'enquêtes sur des processus causals à finalité informationnelle, et non sur des chaînes de calculs symboliques.

Ce déplacement a l'avantage de libérer l'imagination scientifique et philosophique, et de mettre ou de remettre à l'ordre du jour des questions auxquelles le fonctionnalisme strict ne semble pas s'appliquer. Il a l'inconvénient de créer un fossé entre les explications procurées par le cadre du fonctionnalisme strict et les explications processuelles ou causales caractéristiques de la phase présente. Prenons un exemple, celui de la navigation animale : comme y insiste à juste titre Charles Gallistel, le cadre symbolique permet d'expliquer de façon remarquablement économique et fidèle toute une série de comportements animaux, à partir de notions de carte et de calcul sur des variables telles que durée et vitesse d'un parcours. L'incapacité des neurosciences à imaginer la manière dont ces calculs symboliques (sur des *variables* et non des quantités fixées) seraient effectués dans le tissu

neural n'est pas matière à se réjouir : le but n'est pas d'accumuler les arguments contre le fonctionnalisme (même si c'est devenu un sport assez populaire), ni à l'inverse de se gausser de l'« empirisme » des neurosciences (comme dirait Auguste Comte, et comme Fodor aime à le faire). Le but est de découvrir, notamment, comment les insectes trouvent et retrouvent leur chemin. On ne voit pour l'instant guère d'autre stratégie que de rechercher le chaînon manquant, l'articulation entre explication symbolique et explication naturaliste. C'est celle que prône Paul Smolensky, en pensant en particulier à la linguistique, dont une partie considérable est « écrite » dans l'idiome symbolique : faut-il, au prétexte qu'on « ne voit pas comment » un cerveau pourrait réaliser les opérations postulées par ces théories, les abandonner purement et simplement ? Ce serait revenir quelques décennies en arrière : il est plus sensé, et plus conforme aux leçons de l'histoire des sciences, de chercher d'une part des reformulations suffisamment radicales des théories existantes, et d'autre part des hypothèses structurales nouvelles sur le cerveau comme système cognitif naturel.

(c) *La multiplication des écoles, ou l'éclosion de mille fleurs.* L'objectif principal du présent chapitre est de mettre en lumière les nouvelles orientations qui se sont manifestées dans le domaine. Ce choix d'exposition ne doit pas conduire à surévaluer l'importance de ces orientations relativement aux acquis et aux courants issus des périodes antérieures. Il n'en demeure pas moins que l'ébranlement subi par le *mainstream* dans ses fondements et dans sa prétention à être « *the only game in town* » est corrélé avec une explosion de recherches hors *mainstream* et de cadres théoriques différents.

Il serait fastidieux de les passer en revue, et impossible de leur rendre justice dans le cadre de ce chapitre, qui se fait longuet. Parmi les thèmes évoqués dans la première partie, certains se prêtent, ou donnent naissance, à ces cadres non classiques. Une différence importante par rapport à la période antérieure est que l'on s'inquiète beaucoup moins des différences (sauf sans doute dans le cas de la linguistique, où les guerres d'écoles n'ont pas pris fin), et qu'en particulier refont surface sans provoquer de crise des traditions de psychologie scientifique antérieure au tournant cognitif institutionnel, ou contemporaine de ce tournant. Cela va de la perception directe et de la théorie des « affordances » de Gibson (contre laquelle Fodor et Pylyshyn bataillèrent jadis) à un retour de la *Gestalt*, en passant par l'auto-organisation dans les systèmes complexes (appliquée au système nerveux), la reprise de thèmes phénoménologiques (husserliens et merleau-pontiens) dans une perspective naturaliste... Réapparaissent aussi des notions qui, sans avoir été proprement oubliées, avaient pendant un temps semblé sommeiller : la mémoire, l'attention, la douleur... Certaines notions migrent de leur terre d'origine vers les sciences cognitives ; ainsi de la décision, prise en charge par les neurosciences, comme on l'a vu plus haut, mais aussi par une théorie de l'information sensorielle, la théorie de la détection du signal (SDT : *signal detection theory*) qui relève davantage des sciences de l'ingénieur et de l'IA, même si elle s'applique à la psychologie humaine. Des phénomènes longtemps considérés comme peu propices à un travail sérieux sont l'objet d'assauts directs : la folie (les troubles psychotiques), l'acquisition de savoir-faire et d'expertise, les différences individuelles, ou encore la différence entre les sexes sur le plan cognitif. Des hypothèses profondément ancrées et presque universellement partagées sont remises en question : l'idée que les processus mentaux sont composés d'étapes discrètes, accomplies par des modules ou systèmes fonctionnels, déjà malmenée, on l'a vu, par le connexionnisme et par les neurosciences quand elles mettent en évidence des phénomènes d'interaction à distance, est attaquée aussi sur le plan de la théorie psychologique. Le développement

considérable de la modélisation physico-mathématique contribue également à cette « déconstruction » des cadres de pensée hérités de la tradition philosophique et de la psychologie du sens commun.

Cette effervescence a deux conséquences générales. Elle tend d'abord à effacer la distinction entre le *mainstream* et le reste : les sciences cognitives se mettent à ressembler davantage à un marché qu'à une camp militaire. Il est devenu difficile, voire impossible en pratique de discuter du meilleur cadre théorique général, ou de la conformité de telle ou telle démarche à des canons disciplinaires. Les chercheurs, et les philosophes eux-mêmes, semblent préférer s'intéresser aux questions particulières, celles que posent des familles restreintes de phénomènes.

La seconde conséquence est d'effacer, plus rapidement encore, la frontière entre un domaine qui serait celui de la psychologie cognitive et le reste de la psychologie. Comme on l'a dit, la psychologie cognitive, en s'appuyant sur les autres disciplines des sciences cognitives, s'empare d'un thème après l'autre et aura bientôt envahi tout le territoire de la psychologie *théorique*. Dans le même temps, elle se fond elle-même progressivement dans un ensemble disciplinaire plus vaste, et perd ainsi son unité et sa personnalité distinctive. L'étude scientifique de l'esprit, y compris des conditions matérielles de ses manifestations, devient un domaine doté d'une structure spécifique, qui n'est ni celle de la psychologie scientifique des périodes antérieures, ni celle des sciences cognitives des commencements, ni celle d'une science unifiée comme est, ou a la réputation d'être, la physique fondamentale. Il ressemble davantage, comme on l'avait du reste prédit, aux sciences de la vie, qui s'organisent autour de plusieurs « noyaux durs » incomplètement articulés entre eux, et laissent la place à de nombreux programmes de recherche semi-autonomes, en attente paresseuse d'une éventuelle intégration dans une biologie intégralement théorique.

(d) *Formes et limites du naturalisme*. Pour prévisible qu'elle était peut-être, cette situation ne s'en s'écarte pas moins largement des perspectives entretenues par bon nombre de philosophes des sciences cognitives, emportés par un naturalisme intempéré, un optimisme déraisonnable, et trompés selon moi par une vision de la connaissance scientifique que les philosophes des sciences (corporation d'où les plus mûrs d'entre eux sont pourtant issus) ont abondamment critiquée au cours des vingt-cinq dernières années.

Ce que ces philosophes annoncent, c'est une science unifiée de l'esprit susceptible d'achèvement, dans laquelle trouverait place l'explication de toute régularité observée ou observable dans la vie mentale. Or, en premier lieu, la réalisation d'une telle ambition supposerait le comblement de vides immenses sur la carte des terres cognitives connues. Comme y insiste dans ses derniers ouvrages Chomsky, nous ne sommes guère plus avancés que Descartes dans la compréhension de l'application ouverte ou créative de nos facultés. De manière générale, nos progrès dans la description des fonctions supérieures ont été modestes jusqu'à présent. En deuxième lieu, l'unification ne pourrait s'accomplir que dans un cadre conceptuel que nous ne possédons pas : comme on y a insisté, le seul candidat déclaré, le néomécanisme computationnel, ne rallie, ni en théorie ni en pratique, suffisamment de suffrages. En troisième lieu, l'hypothèse naturaliste (l'organe cognitif d'*Homo sapiens* est un système naturel, soumis aux lois de la nature et résultant de l'évolution) ne constitue pas une garantie de succès, même à un horizon infini : en effet, d'une part rien ne garantit que nous ayons les capacités cognitives requises pour déchiffrer les processus en jeu ; d'autre part, rien ne nous justifie dans l'espoir confiant que les contraintes descendantes mise au jour par l'approche fonctionnaliste, même si nous apprenons à l'articuler à la théorie de l'évolution, rencontreront les résultats de l'enquête ascendante conduite sur les mécanismes cérébraux (et plus largement biologiques). En

quatrième lieu, on ne peut inférer d'une régularité dans un phénomène ou un comportement mental l'existence d'un trait du cerveau qui en serait responsable : cette régularité peut être d'origine sociale, culturelle, climatique, écologique, voire volontaire. Certes, ces domaines sont eux-mêmes naturels, en un certain sens, mais le problème de l'aboutissement de la stratégie naturaliste (la rencontre du *top down* et du *bottom up*) se repose à propos des couples individu/société, individu/milieu, etc, exactement dans les termes précédents : en combinant les deux problèmes, les perspectives réalistes de solution s'éloignent encore davantage.

Si nous nous tournons maintenant vers les sciences établies, pour évaluer la situation en la comparant à la leur, nous voyons que l'image de l'unité, de la complétude virtuelle, de la subsomption transparente du particulier sous les lois générales, est contestée par toute une école de philosophie des sciences. Si la réalité est en effet moins conforme à cette image, cela plaide à la fois en faveur d'une réduction des ambitions en matière de cognition, et inversement d'une évaluation à la baisse de l'écart entre les connaissances que nous avons et celles que nous pouvons raisonnablement espérer. Enfin, s'agissant de la vie de l'esprit, un problème inédit se pose : que sont, au juste, les « régularités » de la vie mentale ? De quelle objectivité procèdent-elles ? Ne sont-elles pas « imprégnées » de normativité ? Pour le naturaliste convaincu, cette normativité procède elle-même de la nature, mais, toujours pour les mêmes raisons, cette hypothèse ne garantit pas qu'une solution soit accessible. Cependant, nombre de régularités sont évidemment d'origine directement naturelle, et à ce titre se prêtent à l'enquête naturaliste. Ce sont elles dont les sciences cognitives veulent rendre compte, et cela devrait suffire à les occuper jusqu'à ce que ce volume soit tombé en poussière. Quant à la structure qu'elles acquerront en chemin, elle ne se laisse pas encore deviner, mais elle ne sera sans doute pas celle d'une physique mentale. Elles n'auront pas un grain uniforme, et ressembleront peut-être un peu à la biologie, un peu à l'économie. Faut-il s'en alarmer ?

BIBLIOGRAPHIE

Nota. *A quelques exceptions motivées près, n'ont été retenus que les ouvrages, au détriment des articles, dont on sait pourtant le rôle prépondérant dans la recherche actuelle. Bon nombre d'articles sont désormais disponibles sur les pages web personnelles des auteurs. De même, avec quelques exceptions, ne figurent que des livres parus depuis la préparation de la première édition du présent recueil.*

Allen, C. and Bekoff, M., *Species of minds: The philosophy and biology of cognitive ethology*, Cambridge, MA : MIT Press, 1997

Andler, Daniel, Anne Fagot-Largeault & Bertrand Saint-Sernin, *Philosophie des sciences*, Paris : Gallimard, coll. Folio, 2 vols, 2002 (notamment les chap. 3 et 6 par D. Andler)

Andler, Daniel, *Science et philosophie. Quinze études épistémologiques*, Paris : Bibliothèque du CREA, 1999 (accessible sur Internet)

Arbib, Michael, ed., *The Handbook of Brain Theory and Neural Networks*, Cambridge, MA : MIT Press, 2^e éd., 2003

Atran, Scott, *In Gods We Trust : The evolutionary landscape of religion*, Oxford : Oxford University Press, 2002

Axelrod, Robert, *The Evolution of Cooperation*, New York : Basic Books, 1983

Barkow, Jerome, Leda Cosmides, John Tooby, *The Adapted Mind. Evolutionary Psychology and the Generation of Culture*, New York: Oxford University Press, 1992

Baron-Cohen, Simon *et al.*, eds, *Understanding Other Minds* Oxford : Oxford University Press, 2^e éd., 2000

Baron-Cohen, Simon, *Mindblindness. An Essay on Autism and Theory of Mind*, Cambridge, MA: MIT Press, 1995 ; trad. fr. *La cécité mentale*, Grenoble : Presses de l'université de Grenoble, 1998

Bechtel, William & George Graham, eds., *A Companion to Cognitive Science*, Oxford : Blackwell, 1999

- Bekoff, M., C. Allen, G. Burghardt, *The cognitive animal : Empirical and theoretical perspectives on animal cognition*, Cambridge, MA : MIT Press, 2002
- Bekoff, Marc & Dale Jamieson, eds., *Readings in Animal Cognition*, Cambridge, MA : MIT Press, 1996
- Ben-Ze'ev, Aaron, *The Subtlety of Emotions*, Cambridge, MA : MIT Press, 2000
- Bermúdez, Jose L., Anthony Marcel, & Naomi Eilan, eds., *The Body and the Self*, Cambridge, MA : MIT Press, 1995
- Bermúdez, Jose L., *Thinking without words*, Oxford : Oxford University Press, 2003
- Berthoz, Alain, *Le sens du mouvement*, Paris : Odile Jacob, 1997
- Betzig, L., *Human Nature*, Oxford : Oxford University Press, 1998
- Bickle, J., *Psychoneural Reduction : The new wave*, Cambridge, MA : MIT Press, 1998
- Binmore, K.G., *Game Theory and the Social Contract*, Cambridge, MA : MIT Press, 1994 ; trad. fr. *Jeux et théorie des jeux*, Bruxelles : De Boeck
- Bloch, Maurice, *How We Think They Think: Anthropological Approaches to Cognition, Memory, and Literacy*, Boulder, CO : Westview Press, 1998
- Bloch, Vincent, dir., *Cerveaux et machines*, Paris : Hermès Science, 1999
- Block, Ned, Owen Flanagan, & Güven Güzeldere, eds., *The Nature of Consciousness, Philosophical Debates*, Cambridge, MA : MIT Press, 1997
- Boden, Margaret, ed., *The Philosophy of Artificial Life*, Oxford: Oxford University Press, 1996
- Boudon Raymond, Alban Bouvier, François Chazel, dir., *Cognition et sciences sociales*, Paris : PUF, 1997
- Bourjine, Paul & Jean-Pierre Nadal, *Cognitive Economics , An Interdisciplinary Approach*, New York : Springer, 2004
- Bouvier, Alban, *L'argumentation philosophique, étude de sociologie cognitive*, Paris : PUF, 1995
- Bouvier, Alban, *Philosophie des sciences sociales*, Paris : PUF, 1999
- Bowerman, M., & S. Levinson, *Language acquisition and conceptual development*, Cambridge: Cambridge University Press, 2001
- Boyd, Robert & Joan B. Silk, *How Humans Evolved*, New York : Norton, 3e éd. 2003 (1^e éd. 1997) ; trad. fr. *L'aventure humaine. Des molécules à la culture*, Bruxelles : De Boeck, 2004
- Boyer, Pascal, *Et l'homme créa les Dieux*, Paris : Laffont, 2001
- Boyer, Pascal, *The Naturalness of Religious Ideas: Outline of a cognitive theory of religion*, Los Angeles & Berkeley : University of California Press, 1994
- Bratman Michael, *Faces of Intention*, Cambridge: Cambridge University Press, 1999
- Brooks Rodney A., *Cambrian Intelligence: The Early History of the New AI*, Cambridge, MA : MIT Press, 1999
- Brooks Rodney A., *Flesh and Machine: How robots will change us*, New York : Pantheon Books, 2002
- Brown, M. R. Keynes & A. Lumsden, *The Developing Brain*, Oxford : Oxford University Press, 2001
- Buss, David, *Evolutionary psychology*, Cambridge, MA : Harvard University Press, 1997
- Byrne, R. & A. Whiten, eds., *Machiavellian Intelligence*, Oxford : Oxford University Press, 1988
- Cacioppo, J., G. Berntson, S. Taylor, D. Schacter, *Foundations in social neuroscience*, Cambridge, MA : MIT Press, 2002
- Cacioppo, John et al., eds, *Foundations in Social Neuroscience*, Cambridge, MA : MIT Press, 2002
- Camazine, S. et al., *Self-Organization in Biological Systems*, Princeton : Princeton University Press, 2003
- Campbell, John, *Past, Space, and Self*, Cambridge, MA: MIT Press, 1994
- Carruthers Peter & Andrew Chamberlain, *Evolution and the human mind, Modularity, language and meta-cognition*, Cambridge: Cambridge University Press, 2000
- Carruthers Peter & Peter K. Smith, *Theories of Theories of Mind*, Cambridge: Cambridge University Press, 1996
- Chalmers, David J., *The Conscious Mind. In Search of a Fundamental Theory*, New York: Oxford University Press, 1996
- Chalmers, David, ed., *Philosophy of Mind: Classical and Contemporary Readings*, Oxford : Oxford University Press, 2002
- Changeux, Jean-Pierre, dir., *Gènes et culture*, Paris : Odile Jacob, 2003
- Changeux, Jean-Pierre, *L'homme de vérité*, Paris : Odile Jacob, 2003
- Cheney, D, & R. Seyfarth, *How monkeys see the world*, Chicago : University of Chicago Press, 1990
- Chomsky, Noam, *New Horizons in the Study of Language and Mind*, Cambridge : Cambridge University Press, 2000
- Chomsky, Noam, *On Nature and Language*, Cambridge: Cambridge University Press, 2002

- Chomsky, Noam, *The Minimalist Program*, Cambridge, MA : MIT Press, 1995
- Clapin, Hugh, ed., *Philosophy of Mental Representation*, Oxford : Oxford University Press, 2002
- Clark, Andy, *Being there. Putting Brain, Body, and World Together Again*, Cambridge, MA: MIT Press, 1997
- Cohen, Laurent, *L'homme thermomètre*, Paris : Odile Jacob, 2003
- Cowie, Fiona, *What's within ?*, Oxford : Oxford University Press, 1998
- Crawford, C., & D. Krebs, eds., *Handbook of evolutionary psychology*, Hillsdale, NJ : Erlbaum, 1997
- Crick, Francis & Christof Koch, A framework for consciousness, *Nature Neuroscience* 6 (2), février 2003, pp. 119-126
- Crick, Francis, *The Astonishing Hypothesis: The Scientific Search For The Soul*, New York : Scribner, 1995
- Cummins, Denise & C. Allen, eds., *The Evolution of Mind*, New York : Oxford University Press, 1998
- Cummins, Robert, *Representations, Targets, and Attitudes*, Cambridge, MA : MIT Press, 1996
- D'Andrade, R., *The Development of Cognitive Anthropology*, Cambridge: Cambridge University Press, 1995
- Damasio, Antonio R., *Descartes' Error*, New York: Putnam 1994; trad. fr. *L'erreur de Descartes*, Paris : Odile Jacob, 1995
- Damasio, Antonio R., *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, New York : Harcourt Brace, 1999
- Davidson, Richard J., Klaus R. Scherer, H. Hill Goldsmith, eds., *Handbook of affective sciences*, Oxford : Oxford University Press, 2002
- Dawkins, *The Extended Phenotype*, San Francisco: Freeman, 1982
- Dayan P. & L. Abott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*, Cambridge, MA : MIT Press, 2001
- de Sousa, Ronald, *The Rationality of Emotion*, Cambridge, MA : MIT Press, 1987
- de Waal, F., *Peacemaking among primates*, Cambridge : Cambridge University Press, 1996
- Dehaene, Stanislas, *Le cerveau en action*, Paris : PUF, 1997
- Dehaene, Stanislas, *The Cognitive Neuroscience of Consciousness*, Cambridge, MA : MIT Press, 2001
- DeLancey, Craig, *Passionate Engines: What emotions reveal about the mind and artificial intelligence*, Oxford : Oxford University Press, 2001
- Dennett Daniel C., *Darwin's Dangerous Idea. Evolution and the meanings of life*, New York : Simon & Schuster, 1995; The Penguin Press, 1996; trad. fr. *Darwin est-il dangereux?*, Paris : Odile Jacob, 2002
- Dennett, Daniel C., *Brainchildren: Essays on Designing Minds*, Cambridge, MA : MIT Press, 1998
- Dennett, Daniel C., *Consciousness Explained*, Boston : Little Brown, 1991; trad. fr. *La conscience expliquée*, Paris : Odile Jacob, 1994
- Diamond, Jared, *Guns, Germs, and Steel*, New York : Norton, 1997; trad. fr. *De l'inégalité parmi les sociétés*, Paris : Gallimard, 2000
- Doris, John, *Lack of Character: Personality and Moral Behavior*, Cambridge : Cambridge University Press, 2002
- Dortier, Jean-François, *L'homme, cet étrange animal...*, Auxerre : Sciences Humaines Editions
- Dretske, Fred, *Naturalizing the Mind*, Cambridge, MA : MIT Press, 1995
- Dunbar, Robin, C. Knight & C. Power, *The Evolution of Culture*, Edinburgh : Edinburgh University Press, 1999
- Dunbar, Robin, *Grooming, Gossip, and the Evolution of Language*, Cambridge, MA : Harvard University Press, 1996
- Dupoux, Emmanuel, ed., *Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler*, Cambridge, MA : MIT Press : 2001; trad. fr. *Les langages du cerveau. Textes en l'honneur de Jacques Mehler*, Paris : Odile Jacob, 2002
- Dupuy, Jean-Pierre, *Aux origines des sciences cognitives*, Paris : La Découverte, 1999
- Edelman, Gerald M. & Giulio Tononi, *A Universe of Consciousness: How Matter Becomes Imagination*, New York : Basic Books, 2001
- Ekman, P., & R.J. Davidson, *The nature of emotions, Fundamental questions*, Oxford : Oxford University Press, 1994
- Eliassmith, Ch., & C. Anderson, *Neural engineering : computation, representation, and dynamics in neurobiological systems*, Cambridge, MA : MIT Press, 2002
- Elman Jeffrey, L. E. Bates, M.H. Johnson, A. Karmiloff-Smith, D. Parisi, & D. Plunkett, *Rethinking Innateness. A Connectionist Perspective on Development*, Cambridge, MA : MIT Press 1996
- Elster, Jon, *Alchemies of the Mind: Rationality and the emotions*, Cambridge: Cambridge University Press, 1999; trad. fr. partielle *Proverbes, maximes, émotions*, Paris : PUF, 2003
- Engel, Pascal, *Philosophie et psychologie*, Paris : Gallimard, coll. Folio, 1996
- Fagin, Ronald, Joseph Y. Halpern, Yoram Moses & Moshe Y. Vardi, *Reasoning About Knowledge*, Cambridge, MA : MIT Press, 2e éd., 2003
- Fehr, Ernst & Armin Falk, Psychological foundations of incentives, *European Economic Review* 46 (2002), pp. 687-724

- Fisette, D. & P. Poirier, dir., *Philosophie de l'esprit*, Paris : Vrin, 2 volumes, 2002-3
- Fodor, Jerry, *Concepts. Where cognitive science went wrong*, Oxford : Oxford University Press, 1998
- Fodor, Jerry, *The mind doesn't work that way. The scope and limits of computational psychology*, Cambridge, MA : MIT Press, 2000; trad. fr. *L'esprit, ça ne marche pas comme ça*, Paris : Odile Jacob, 2003
- Forrest, Stephanie, ed., *Emergent Computation*, Cambridge, MA : MIT Press 1991
- Frank, Robert, *Passions Within Reason: The strategic role of emotions*, New York : Norton, 1988
- Frith, Christopher, *The Cognitive Neuropsychology of Schizophrenia*, Hillsdale: Lawrence Erlbaum. 1992; trad. fr. *Neuropsychologie cognitive de la schizophrénie*, Paris : PUF, 1996
- Frith, Uta, *Autism: Explaining the Enigma*, Oxford : Blackwell Publishers, 2e éd. 2003
- Gallistel, Charles, *The Organization of Learning*, Cambridge, MA : MIT Press, 1990
- Gazzaniga, Michael S., ed., *The New Cognitive Neurosciences: Second Edition*, Cambridge, MA : MIT Press, 1999
- Gentner, Dedre & Susan Goldin-Meadow, *Language in Mind*, Cambridge, MA : MIT Press, 2003
- Gibbard, Allan, *Wise Choices, Apt Feelings: A theory of normative judgment*, Oxford : Oxford University Press, 1990; trad. fr. *Sagesse des choix, justesse des sentiments*, Paris : PUF, 1996
- Gigerenzer, Gerd & R. Selten, *Bounded rationality*, Cambridge, MA : MIT Press, 2002
- Gigerenzer, Gerd and the ABC group, *Simple heuristics that make us smart*, Oxford : Oxford University Press, 1999
- Gigerenzer, Gerd, *Adaptive Thinking. Rationality in the real world*, Oxford : Oxford University Press, 2000
- Gilbert, D., S. Fiske, G. Lindzey, F. Lindzey, eds., *The handbook of social cognition*, Oxford : Oxford University Press, 1998
- Gilbert, Margaret, *Marcher ensemble. Essais sur le fondements des phénomènes collectifs*, Paris : PUF, 2003
- Gilbert, Margaret, *On Social Facts*, Londres : Routledge, 1989, réimp. Princeton : Princeton University Press, 1992
- Gilbert, Margaret, *Sociality and Responsibility*, Lanham, MD : Rowman & Littlefield, 2000
- Gilovitch, Griffin & Daniel Kahnemann, eds., *Heuristics and biases: The psychology of intuitive judgment*, Cambridge: Cambridge University Press, 2002
- Gintis, H., *Game theory evolving*, Princeton : Princeton University Press, 2000
- Glimcher, P., *Decisions, Uncertainty, and the Brain*, Cambridge, MA : MIT Press, 2003
- Godfrey-Smith, Peter, *Complexity and the Function of Mind in Nature*, Cambridge: Cambridge University Press, 1998
- Goldstone, Robert L. & Lawrence W. Barsalou, "Reuniting perception and conception", *Cognition* 65, 1998, pp. 231-262
- Goody, Jack, *Entre l'oralité et l'écriture*, Paris : PUF, 1994
- Goody, Jack, *The Logic of Writing and the Organization of Society*, Cambridge : Cambridge University Press, 1987
- Griffin, Donald R., *Animal Minds*, Chicago : University of Chicago Press, 1992
- Griffin, Donald, *The Question of Animal Awareness*, New York : Rockefeller University Press, 1976; 2e éd. augm. et rév., 1981
- Griffiths, Paul, *What Emotions Really Are: The problem of psychological categories*, Chicago : University of Chicago Press, 1997
- Gunther, York H., ed., *Essays on Nonconceptual Content*, Cambridge, MA : MIT Press, 2003
- Guttenplan, Samuel, ed., *A Companion to the Philosophy of Mind*, Oxford : Blackwell, 1996
- Hacking, Ian, *Rewriting the Soul: Multiple Personality and the Sciences of Memory*, Princeton : Princeton University Press, 1995; trad. fr. *L'âme réécrite*, Paris : Les empêcheurs de penser en rond, 1998
- Hacking, Ian, *The Social Construction of What?*, Cambridge, MA : Harvard University Press, 1999
- Hauser, Mark, *A quoi pensent les animaux ?*, Paris : Odile Jacob, 2002
- Hauser, Mark, Noam Chomsky, W. Tecumseh Fitch, The faculty of language: What is it, who has it, and how did it evolve?, *Science* **298**, 22 nov. 2002
- Hirschfeld, Lawrence A. & Susan A. Gelman, eds., *Mapping the Mind. Domain Specificity in Cognition and Culture*, Cambridge: Cambridge University Press, 1994
- Hirschfeld, Lawrence, *Race in the Making: Cognition, Culture, and the Child's Construction of Human Kinds*, Cambridge, MA : MIT Press, 1996.
- Houdé, Olivier et al., *Vocabulaire de sciences cognitives*, Paris : PUF, 1998
- Houdé, Olivier, Bernard Mazoyer, Nathalie Tzourio-Mazoyer et al., *Cerveau et psychologie*, Paris : PUF, 2002
- Hurley, S., *Consciousness in Action*, Cambridge, MA : MIT Press, 1998
- Hutchins, Edwin, *Cognition in the Wild*, Cambridge, MA: MIT Press, 1995

- Jackendoff, Ray, *Foundations of Language*, Oxford : Oxford University Press, 2002
- Jackendoff, Ray, *Languages of the Mind: Essays on Mental Representation*, Cambridge, MA : MIT Press, 1992
- Jacob, Pierre, Marc Jeannerod, Charles Heckscher, *Ways of Seeing: The Scope and Limits of Visual Cognition*, Oxford : Oxford University Press, 2003
- Jacob, Pierre, *Pourquoi les choses ont-elles un sens?*, Paris : Odile Jacob, 1997
- Jeannerod, Marc, *The Cognitive Neuroscience of Action*, Cambridge: Cambridge University Press, 1997
- Johnson D.M. & C.E. Erneling, eds., *The Future of the Cognitive Revolution*, New York: Oxford Univ. Press 1997
- Juarrero, A., *Dynamics in Action*, Cambridge, MA : MIT Press, 1999
- Kandel, Eric R., James H. Schwartz, Thomas M. Jessel, eds., *Principles of Neural Science*, New York : McGraw Hill, 4th ed., 2000
- Kanizsa, Gaetano, *Grammatica del vedere*, Il Mulino ; trad. fr. *La grammaire du voir*, Paris : Diderot, 1997
- Karmiloff-Smith, Annette, *Beyond Modularity, A Developmental Perspective on Cognitive Science*, Cambridge, MA: MIT Press 1992
- Kauffman, Stuart, *At Home in the Universe: The Search for Laws of Self-Organization and Complexity*, Oxford : Oxford University Press, 1996
- Kaufmann, Stuart, *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford : Oxford University Press, 1993
- Kayser, Daniel, *La représentation des connaissances*, Paris : Hermès Science, 1997
- Keil, F., & R. Wilson, R., *Explanation and Cognition*, Cambridge, MA : MIT Press, 2000
- Kim, Jaegwon, *Mind in a physical world*, Cambridge, MA : MIT Press, 1998
- Kimura, Doreen, *Sex and Cognition*, Cambridge, MA : MIT Press, 1999
- Klein, R., *The Human Career: Human biological and cultural origins*, 2e éd., Chicago : University of Chicago Press, 1999
- Koch, Christof, *The Quest for Consciousness: A neurobiological approach*, Englewood, CO : Roberts & Co, 2004
- Krebs, J R, & N. Davies, *Behavioral Ecology : An evolutionary approach*, Oxford : Blackwell, 1997
- Kunda, Z., *Social Cognition*, Cambridge, MA : MIT Press, 1999
- Laland, K., N., & G. Brown, *Sense and nonsense : Evolutionary perspectives on human behavior*, Oxford : Oxford University Press, 2002
- Lambie, John A. & Anthony J. Marcel, Consciousness and the varieties of emotion experience: A theoretical framework, *Psychological Review* **109** 2 (2002), pp. 219-259
- Lane, R.D., & Lynn Nadel, eds., *Cognitive Neuroscience of Emotion*, Oxford : Oxford University Press, 2000
- Lazarus, R.S. & B.N. Lazarus, *Passion and reason: Making sense of our emotions*, Oxford : Oxford University Press, 1994
- LeDoux, Joseph, *The Emotional Brain: The mysterious underpinnings of emotional life*, New York : Simon and Schuster, 1996
- Levine, Joseph, *Purple Haze*, Oxford : Oxford University Press, 2001
- Levinson, S., ed., *Evolution and Culture*, Cambridge, MA : MIT Press, 2004
- Lewis, Mark & Jeannette M. Haviland-Jones, *Handbook of Emotions*, New York : Guilford Press, 2^e éd. 2000
- Libet, Benjamin, *Neurophysiology of Consciousness: Selected Papers and New Essays*, Boston : Birkhäuser, 1995
- Livet, Pierre, dir., *De la perception à l'action*, Paris :Vrin, 2000
- Livet, Pierre, dir., *Révision des croyances*, Paris : Hermès Science, 2002
- Livet, Pierre, *Emotions et rationalité morale*, Paris : PUF, 2002
- Lycan, William, ed., *Mind and Cognition: An Anthology*, Oxford : Blackwell, 2e éd. 1999
- Macnamara, John & Gonzalo Reyes, eds., *Logical Foundations of Cognition*, New York: Oxford Univ. Press, 1994
- Malle, Bertram F., Louis J. & Dare A. Baldwin, eds., *Intentions and Intentionality : Foundations of Social Cognition*, Cambridge, MA : MIT Press 2003
- Margolis, Eric & Stephen Laurence, eds., *Concepts. Core readings*, Cambridge, MA : MIT Press, 1999
- May, Larry, Marilyn Friedman & Andy Clark, eds., *Mind and Morals*, Cambridge, MA: MIT Press 1996
- McAdams, Stephen, Emmanuel Bigand et collab., *Penser les sons. Psychologie cognitive de l'audition*, Paris : PUF, 1994
- McClamrock, R., *Existential Cognition*. Chicago : University of Chicago Press, 1995
- McDowell, John, *Mind and World*, Cambridge, MA : Harvard University Press, 1994
- McGinn, Colin, *The Problem of Consciousness*, Oxford : Blackwell, 1991
- Medin, Douglas & Scott Atran , eds., *Folk Biology*, Cambridge, MA : MIT Press, 1998

- Milner, David & Melvyn Goodale, *The Visual Brain in Action*, Oxford : Oxford University Press, 1995
- Mithen, Steven, *The Prehistory of the Mind*, Londres : Thames & Hudson, 1996
- Murphy, Gregory L., *The Big Book of Concepts*, Cambridge, MA : MIT Press, 2002
- Nadel Jacqueline & Jean Decety, *Imiter pour découvrir l'humain*, Paris : PUF, 2002
- Nadel, Lynn, gen. ed., *Encyclopaedia of Cognitive Science*, Londres, new York, Tokyo : Nature Publishing Group, 4 vols., 2003
- Nagel, Thomas, *Mortal Questions*, Cambridge: Cambridge University Press, 1979 (contient l'article de 1974, "What is it like to be a bat?"); trad. fr. *Questions mortelles*, Paris : PUF
- NBIC : v. Roco & Bainbridge
- Neisser, Ulric, ed., *The Perceived Self: Ecological and Interpersonal Sources of Self-Knowledge*, New York : Cambridge University Press, 1993
- Nelkin, Norton, *Consciousness and the Origins of Thought*, Cambridge: Cambridge Univ. Press 1996
- Neu, Jerome, *A Tear is an Intellectual Thing: The meanings of emotions*, Oxford : Oxford University Press, 2000
- Nisbett, Richard & D. Cohen, *The culture of honor*, Boulder, CO : Westview Press, 1996
- Núñez, Rafael & Walter Freeman, eds., *Reclaiming Cognition. The primacy of action, intention and emotion*, Thoreverton, R.-U. : Imprint Academic, 1999
- Oatley, K. & J.M. Jenkins, *Understanding Emotions*, Oxford : Blackwell, 1996
- Osherson, Daniel, gen. ed., *An Invitation to Cognitive Science*, Cambridge, MA : MIT Press, 4 volumes, 2e éd. 1995-1998
- Panksepp, Jaak, *Affective Neuroscience: The foundations of human and animal emotions*, New York, Oxford : Oxford University Press, 1998
- Parfit, Derek, *Reasons and Persons*, Oxford : Oxford University Press, 1984
- Penrose, Roger, *Shadows of the Mind: An approach to the missing science of consciousness*, Oxford : Oxford University Press, 1994; trad. fr. •••?
- Perry, John, *Reference and Reflexivity*, Stanford, CA : CSLI Publications, 2001
- Perry, John, *The Problem of the Essential Indexical and Other Essays*, New York: Oxford Univ. Press, 1993
- Petitot, Jean, Francisco Varela, Bernard Pachoud, Jean-Michel Roy, dir., *Naturaliser la phénoménologie. Essais sur la phénoménologie contemporaine et les sciences cognitives*, Paris : CNRS Editions, 2002
- Pinker Steven, *The Language Instinct. How the Mind Creates Language*, New York : William Morrow, 1994; éd. de poche New York : Harper-Collins, 1995; trad. fr. *L'instinct du langage*, Paris : Odile Jacob, 1999
- Politzer, Guy, dir., *Le raisonnement humain*, Paris : Hermès, 2002
- Port, Robert F. & Timothy van Gelder, eds., *Mind as Motion: Dynamical perspectives on behavior and cognition*, Cambridge, MA: MIT Press, 1995
- Povinelli, D., *Folk physics for apes*, Oxford : Oxford University Press, 2003
- Premack, David & Ann J. Premack, *Original Intelligence: The Architecture of the Human Mind*, New York : McGraw Hill, 2002 ; trad. fr. *Le bébé, le singe et l'homme*, Paris : Odile Jacob, 2003
- Prince, Alan & Paul Smolensky, *Optimality Theory: Constraint interaction in generative grammar*, Oxford & New York : . Blackwell, sous presse. (Disponible depuis 1993 sous la forme de Technical Report CU-CS-696-93, Department of Computer Science, University of Colorado at Boulder, et Technical Report TR-2, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ.
- Prinz, Jessie, *Furnishing the Mind: Concepts and Their Perceptual Basis*, Cambridge, MA: MIT Press, 2002
- Pylyshyn, Z., *Seeing and visualizing : It's not what you think*, Cambridge, MA : MIT Press, 2003
- Quartz, S.R., and Sejnowski, T.J., *Liars, Lovers, and Heroes: What the New Brain Science Reveals About How We Become Who We Are*, New York : HarperCollins, 2002
- Quartz, Steven R. & Terrence J. Sejnowski, "The neural basis of cognitive development: A constructivist manifesto", *Behavioral and Brain Sciences* 20 4, Dec. 1997, pp. 537-596
- Récanati, François, *Direct Reference: From Language to Thought*, Oxford : Blackwell, 1997
- Richerson, Peter J. & Robert Boyd, *The Nature of Culture and Cultures*. sous presse
- Roberts, William A., *Principles of Animal Cognition*, Englewood Cliffs, NJ : Prentice-Hall, 1998
- Roco, M.C., W.S. Bainbridge, eds., "Converging technologies for improving human performance: Nano-, Bio-, Info-Technology and Cognitive Science", National Science Foundation and Department of Commerce Report, 2002; <http://wttec.org/ConvergingTechnologies/> ou <http://www.nsf.gov/nano>
- Rolls, E.T. & G. Deco, *Computational Neuroscience of Vision*, Oxford : Oxford University Press, 2002
- Ross, L. & R. Nisbett, *The person and the situation*, New York : McGraw Hill, 1991
- Rowlands, M., *The Body in Mind*, Cambridge Cambridge University Press, 1999

- Russell, Stuart, and Peter Norvig, *Artificial Intelligence*, Englewood Cliffs, NJ : Prentice-Hall, 1995, 2^e éd. 2003
- Savage-Rumbaugh, E. Sue, Stuart Shanker, Talbot J. Taylor, *Apes, Language, and the Human Mind*, Oxford : Oxford University Press, 1998
- Searle, John, *The Construction of Social Reality*, New York : The Free Press, 1995
- Shallice Tim, *From Neuropsychology to Mental Structure*, Cambridge: Cambridge University Press, 1988; trad. fr. *Symptômes et modèles en neuropsychologie*, Paris : PUF, 1995
- Simon, Herbert A., *Models of Bounded rationality*, Cambridge, MA : MIT Press, 1982
- Skyrms, Brian, *Evolution of the Social Contract*, Cambridge: Cambridge University Press, 1996
- Smith, Brian Cantwell, *On the Origin of Objects*, Cambridge, MA : MIT Press, 1996
- Smolensky, Paul & Géraldine Legendre, *The harmonic mind: From neural computation to optimality-theoretic grammar*, vol. 1: Cognitive Architecture; vol. 2: Linguistic and Philosophical Implications, Cambridge, MA : MIT Press, 2004
- Snyder, C.R. & S.J. Lopez, *Handbook of positive psychology*, Oxford : Oxford University Press, 2002
- Sober, Elliott and Wilson, David, *Unto others: The evolution and psychology of unselfish behavior*, Cambridge, MA : Harvard University Press, 1998
- Solomon, Robert C., ed., *Thinking About Feelings: Philosophers on emotion*, Oxford : Oxford University Press, 2003
- Sperber, Dan, David Premack, & Ann James Premack, eds., *Causal Cognition. A multidisciplinary debate*, Oxford : Clarendon Press, 1995
- Sperber, Dan, ed., *Metarepresentations: A Multidisciplinary Perspective*, Oxford : Oxford University Press, 2000
- Sperber, Dan, *La contagion des idées*, Paris : Odile Jacob, 1996
- Stalnaker, Robert C., *Context and Content*, Oxford : Oxford University Press, 1999
- Stemenov, Maksim I. & Vittorio Gallese, *Mirror Neurons and the Evolution of Brain and Language*, Amsterdam & Philadelphie : John Benjamins, 2002
- Stephens, G. Lynn & George Graham, *When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts*, Cambridge, MA : MIT Press, 2000
- Sterelny, Kim, *Thought in an Hostile World*, Oxford : Blackwell, 2003
- Stevens, Anthony, *Evolutionary Psychiatry : A new beginning*, Oxford : Routledge, 2^e éd., 2000
- Stich, Stephen P., *Deconstructing the Mind*, New York: Oxford Univ. Press 1996
- Stocker, Michael, with Elizabeth Hegeman, *Valuing Emotions*, Cambridge: Cambridge University Press, 1992
- Swets, John A., *Signal Detection Theory and ROC analysis in psychology and diagnostics*, Matwah, NJ : Erlbaum, 1996
- Tesar, B. & P. Smolensky, *Learnability in Optimality Theory*, Cambridge, MA : MIT Press, 2000
- Thelen, Esther & Linda B. Smith, *A Dynamic System Approach to the Development of Cognition and Action*, Cambridge, MA: MIT Press 1994
- Tolman, Edward C., Cognitive maps in rats and men, *Psychological Review* **55** (1948), pp. 189-208
- Tomasello, Michael & Josep Call, *Primate Cognition*, Oxford : Oxford University Press, 1997
- Tomasello, Michael, *The Cultural Origins of Human Cognition*, Cambridge, MA : Harvard University Press, 2001
- Trivers, R., *Natural Selection and Social Theory*, Oxford : Oxford University Press, 2002
- Tye, Michael, *Ten Problems of Consciousness. A Representational Theory of the Phenomenal Mind*, Cambridge, MA : MIT Press, 1995
- Uexküll, Jacob von, *Streifzüge durch die Umwelten von Tieren und Menschen – Bedeutungslehre*, Hamburg: Rowohlt 1956; trad. fr. Ph. Muller, *Mondes animaux et mond humain, suivi de Théorie de la signification*, Paris: Gonthier 1965
- Vapnik, V. N. *The nature of statistical learning theory*, Berlin : Springer, 1995
- Vauclair, Jacques, *L'homme et le singe : psychologie comparée*, Paris : Flammarion, 1998
- Visetti, Yves-Marie & Victor Rosenthal, *Köhler*, Paris : Les Belles-Lettres, 2003
- Walliser Bernard, *L'économie cognitive*, Paris : Odile Jacob, 2000
- Walsh, Vincent & Alvaro Pascual-Leone, *Transcranial Magnetic Stimulation : A Neurochronometrics of Mind*, Cambridge, MA : MIT Press, 2003
- Weiskrantz, L., *Consciousness lost and found*, Oxford : Oxford University Press, 1997
- Whiten, A., and Byrne, R, *Machiavelian intelligence II*, Cambridge: Cambridge University Press, 1997
- Wilson, R., *Cartesian Psychology and Physical Minds*, Cambridge : Cambridge University Press, 1995
- Wilson, Robert A. & Frank C. Keil, eds., *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge, MA : MIT Press, 2001 [MITECS]
- Young, H.P., *Individual strategy and social structure: An evolutionary theory of institutions*, Princeton : Princeton University Press, 1998

Zerubavel, Eviatar, *Social Mindscapes. An invitation to cognitive sociology*, Cambridge, MA : Harvard University Press, 2003