

ENTRE NEUROSCIENCES & PSYCHOLOGIE COGNITIVE : UNE FRONTIÈRE EN QUESTION

Alain TÊTE

RÉSUMÉ : *L'apparition des modèles connexionnistes dans les années 1980 a transformé le problème de la frontière séparant les neurosciences de la psychologie cognitive. Alors que les modèles cognitivistes de traitement symbolique s'inspiraient directement de « l'architecture von Neumann » des ordinateurs et laissaient aux neurosciences le soin de décrire en termes physicalistes l'inscription matérielle du symbolique (« l'implémentation »), les modèles connexionnistes proposent des formalismes mathématiques qui rendent compte, en termes de systèmes dynamiques, de cette implémentation. À la frontière entre niveaux neuronaux et niveaux symboliques se substitue une lecture nouvelle des uns et des autres.*

Il arrive, dans l'histoire des sciences, qu'un événement modifie de façon durable les relations — donc les frontières — entre des disciplines voisines. Ce peut être la découverte d'une réalité jusqu'alors insoupçonnée (la spirale de l'ADN), la mise en œuvre d'une nouvelle technique (la microscopie électronique) ou d'un formalisme (le connexionnisme). Cet événement remet alors en cause les limites traditionnelles de domaines scientifiques mitoyens, soit qu'il en précise les linéaments, soit qu'il *annule leur mitoyenneté* en promouvant une nouvelle discipline qui, s'interposant entre ces territoires séculaires, en transforme profondément les rapports.

C'est ainsi qu'entre les neurosciences et la psychologie cognitive un événement épistémologique est survenu qui a modifié en quelques années la problématique de leurs relations. Jusqu'en 1986, leurs frontières traversaient des *terrae incognitae* suffisamment vastes pour que leurs discours respectifs préservent leur souveraineté. Il était convenu que la « machinerie cognitive » simulée par l'ordinateur et la « machinerie nerveuse » étaient des champs d'étude spécifiques, utilisant des langages distincts mais indispensables l'un et l'autre pour circonscrire le problème de l'union de l'esprit et du corps.

Avec la publication en 1986 de *Parallel Distributed Processing : Explorations in the Microstructure of Cognition*, D. E. Rumelhart, McClelland et le PDP Research Group du M.I.T. bouleversent les données du problème

corps-esprit (dans la littérature anglo-saxonne, *mind-body problem*) en utilisant dans cette « bible » un nouveau langage formel, le connexionnisme.

Le choix de cette cheville historique pour décrire l'état des lieux peut être contesté (les premiers travaux néo-connexionnistes apparaissent en effet à la fin des années 1970)¹, mais il reste que les chercheurs — qu'ils soient neuroscientifiques ou cognitivistes — ne formulent plus, après 1986, le problème de leurs frontières de la même façon qu'auparavant. Le connexionnisme se présente bien comme un « événement épistémologique » : ici se précisent l'état de la question au moment où il surgit et les nouvelles données du problème.

1. — LE CERVEAU CALCULATEUR

Qu'en était-il donc au début des années 80 ?

Deux ensembles de disciplines, les neurosciences et les sciences cognitives se répartissent alors l'étude des processus nerveux et des processus cognitifs. Ces ensembles débordent depuis une vingtaine d'années les champs traditionnels de la neurophysiologie et de la psychologie expérimentale entre lesquelles la psychophysiologie avait occupé une place frontalière au statut ambigu.

Neurosciences et sciences cognitives se partagent des observables (physiques) et des inobservables (mentaux), hiérarchiquement distribués en *niveaux d'organisation* auxquels correspondent, entre les deux extrêmes — biophysique et psychosociologique — des disciplines parmi lesquelles la biologie moléculaire, la neurophysiologie et la psychologie cognitive. Chacune de ces disciplines se donne pour objet d'intégrer dans un même système explicatif les phénomènes relevant de son niveau. Cette division des domaines de recherche se fonde sur une hiérarchie de processus selon l'espace et le temps, mais aussi selon leur caractère structural ou fonctionnel. Elle en restitue la complexité à proportion du « grain » des entités ainsi abstraites². En principe, une telle distribution par niveaux d'organisation devrait garantir la *complémentarité* institutionnelle des disciplines puisqu'elle s'appuie sur la cohérence interne de chaque niveau de réalité et qu'elle postule entre ces niveaux des transitions rationnelles.

1. William BECHTEL, Adele ABRAHAMSEN, *Connectionism and the Mind*, Oxford, Basil Blackwell, 1991, trad. fr. *Le Connexionnisme et l'esprit*, Paris, La Découverte, 1993, p. 25.

2. Jean REQUIN, « Les neurosciences cognitives : au-delà du réductionnisme, une science de synthèse? », 1985, in *Comportement, cognition, conscience*, Paris, Presses universitaires de France, 1987, p. 38.

Cette répartition par niveaux suscite chez les psychologues cognitivistes un certain malaise théorique. Leur position est de plus en plus inconfortable, tiraillés qu'ils sont par leurs liens avec l'Intelligence Artificielle et par ceux qu'ils ne peuvent rompre avec les neurosciences.

D'un côté, ils ne sauraient renoncer à la simulation computationnelle des comportements cognitifs. Les systèmes symboliques matériels (SSM) dans lesquels s'incarne le calcul ont acquis, avec la machine universelle de Turing (1937) puis avec les ordinateurs, un statut épistémologique fondamental puisque ces systèmes symboliques délogent la pensée logique de l'esprit humain qui en était jusqu'ici le lieu exclusif et l'incarnent dans des systèmes artificiels (des modèles) totalement rationalisés.

Mais d'un autre côté, les psychologues ne peuvent oublier que les opérations cognitives chez les animaux et l'homme mobilisent des structures neurologiques qui contrastent de façon saisissante avec celle des ordinateurs « d'architecture von Neumann »³ qui leur servent de modèles. Si les calculateurs rivalisent avec l'activité hautement rationnelle dont le cerveau humain est capable, l'architecture de celui-ci présente une complexité sans commune mesure avec celle de ces machines.

S'ensuit une répartition ingénieuse et efficace des tâches entre psychologues et neuroscientifiques fondée sur la distinction de la fonction et de la structure des systèmes cognitifs. Les équivalences fonctionnelles, qui sont supposées par la modélisation sur ordinateurs des processus cognitifs, ne prennent pas en compte les structures matérielles (que ce soient des puces de silicium ou des neurones). Celles-ci sont posées comme autant de « boîtes noires » dont les propriétés retenues satisfont aux conditions minimales du fonctionnement computationnel. On appelle « implémentation » cette matérialisation dans un ordinateur ou dans le cerveau. Le réseau de neurones formels de McCulloch et Pitts (1943)⁴ ne retient du neurone « humide » que la modalité électrique de sa transmission synaptique (au moins deux entrées, l'une excitatrice, l'autre inhibitrice, une sortie et un seuil avec traitement binaire de l'information) et néglige délibérément le plus clair des connaissances de l'époque portant sur la structure anatomophysiological du cortex (réseaux non linéaires, complexité des connexions, activité humorale, etc.). C'est à ce prix que les cybernéticiens — bien que très informés en matière de neurologie — vont faire de l'ordinateur le médiateur paradigmatique des sciences cognitives. « Nous traitons le cerveau comme une machine de Turing, c'est-à-dire comme un dispositif qui

3. Jacques PAILLARD, « L'ordinateur et le cerveau : un contraste saisissant », *Afset/Interface*, 57, juil. 1987, p. 4.

4. Warren S. MCCULLOCH, Walter PITTS, « A Logical Calculus of the Ideas Immanent in Nervous Activity », *Bulletin of Mathematical Biophysics*, 5, 1943, p. 115-133, trad. fr. in *Sciences cognitives. Textes fondateurs*, Paris, Presses universitaires de France, 1995.

puisse remplir le genre de fonctions qu'un cerveau doit accomplir », dira plus tard McCulloch⁵ évoquant son fameux article de 1943.

« Traiter le cerveau comme une machine de Turing » excède les limites d'une simple analogie. C'est d'une équivalence *stricte* qu'il s'agit, fondée sur la démonstration turingienne que tout ce qui pouvait être calculé par un homme pouvait également l'être par une machine⁶. Ce mécanisme théorique — qui ne doit rien à l'état du machinisme de son temps — contraste avec le mécanisme cartésien dont la rationalité imite et prolonge la rationalité de la mécanique de la Nature : « Toutes les règles des mécaniques appartiennent à la physique, dit Descartes, de sorte que toutes les choses qui sont artificielles sont avec cela naturelles »⁷. La machine de Turing, en revanche, appartient à la physique parce qu'elle appartient à la pensée calculatrice. Théorique, sa structure est intégralement dépendante de sa *fonction* qui consiste à décider, pour toute assertion qui lui serait présentée, si elle est démontrable ou non. Machine logique et non biologique, sa raison d'être n'est pas d'imiter le comportement d'un vivant mais de calculer tout ce qu'une pensée est susceptible de calculer.

La relation d'équivalence entre « machine universelle » de Turing et pensée rationnelle va complètement transformer la classique problématique du corps et de l'esprit. Tandis que l'approche biologique inscrit le corps humain dans le continuum évolutif des espèces animales et confère au système nerveux central une fonction *adaptative* grâce à quoi le vivant (animal ou humain) survit dans un environnement, la « machine de Turing » ignore cette fonction adaptative car elle n'est pas équivalente au corps humain mais à la pensée la plus abstraite. Si tout calcul suppose une « mise en mots » et porte donc sur des symboles, c'est bien la substance pensante de Descartes, reconnaissable à son langage, qui se trouve ainsi concernée dans la simulation turingienne. Par le mécanisme qu'il invoquait dans son *Traité de l'Homme*, Descartes, à l'opposé, voulait rendre compte non de l'esprit mais du corps, cette « machine de terre » comme il l'appelle. L'homme qui, par son corps, était pour Descartes un animal-machine devient avec Turing une pensée-machine dont la corporité *n'excède pas celle du cortex quand il opère sur des représentations symboliques*.

5. W. S. McCULLOCH, dans la discussion qui suit la conférence de J. von Neumann au Hixon Symposium (1948), « The General and Logical Theory of Automata », in *Cerebral Mechanism in Behavior*, L. A. JEFFRIES, ed., New York, John Wiley & Sons Inc., 1951.

6. Pour une description cursive de la « machine de Turing » d'un accès aisé, consulter Daniel ANDLER, *Introduction aux sciences cognitives*, Paris, Gallimard, « Folio », 1992, p. 22. Sur l'œuvre et la vie d'Alan Turing, lire Andrew HODGES, *Alan Turing : the Enigma of Intelligence*, 1983, trad. fr. *Alan Turing ou l'énigme de l'intelligence*, Paris, Payot, « Bibliothèque historique Payot », 1988.

7. René DESCARTES, *Principia philosophiae*, 1644, art. 203, in *Œuvres*, Paris, Gallimard, « Bibliothèque de la Pléiade », 1953, p. 666.

Conséquence directe de cette « aspiration vers le haut » du rapport entre le corps calculateur de la machine et la pensée logico-mathématique : la psychologie cognitive peut se permettre de négliger les recherches des neurosciences dès l'instant qu'elle se donne un cerveau *ad hoc*, celui des calculateurs. Devenue psychologie de l'esprit alors que le behaviorisme la voulait psychologie du comportement (animal ou humain), la psychologie perd dans le domaine des fonctions adaptatives ce qu'elle gagne dans le domaine des fonctions symboliques. Là où les neurosciences pensent en termes biophysiques, elle pense en termes de résolution de problèmes ou d'heuristiques. À la relation adaptative à l'environnement, elle substitue un traitement computationnel d'informations au moyen d'algorithmes. Tout entière investie dans la simulation des processus cognitifs logico-mathématiques, la psychologie cognitive se donne pour programme de *descendre* du logique au biologique (en jargon *top-down*) tandis que les neurosciences se proposent de *monter* du biologique au logique (en jargon *bottom-up*).

2. — HIÉRARCHISATION ASCENDANTE OU DESCENDANTE ?

Une double hiérarchisation des mêmes processus apparaît alors comme l'effet d'une double exigence. D'une part, celle d'un réductionnisme ascendant, celui des neurosciences, qui cherche à rendre compte, à partir des structures neuronales de la fonction cognitive. De l'autre, l'exigence d'un réductionnisme descendant, celui de la psychologie cognitive (appelée encore « computo-représentationnelle » puisque portant sur du calculable et du symbolique), qui vise à aboutir à des structures neuronales compatibles avec les fonctions cognitives. Comme le résume Jean Requin :

« Ni la valeur explicative ou même heuristique, d'une neurobiologie réduisant les phénomènes qui sont l'objet de la psychologie cognitive, ni celle d'une psychologie cognitive réduisant les phénomènes qu'étudie le neurobiologiste ne seraient *a priori* exclues, sans être pour autant érigées en un principe exclusif de l'explication scientifique. En d'autres termes, à l'ordre que peuvent légitimement prétendre mettre les systématisations de l'organisation neuronale dans la compréhension des faits de comportement, répondrait l'ordre que peuvent tout aussi légitimement prétendre mettre les modélisations des phénomènes cognitifs dans la compréhension des fonctions nerveuses »⁸.

Ces « mises en ordre » respectives des neurosciences et des sciences de la cognition constituent un programme dont la vertu œcuménique est mani-

8. J. REQUIN, *op. cit. supra* n. 2, p. 41.

festes : éviter qu'un réductionnisme d'inspiration physicaliste nuise au développement des recherches inspirées par la modélisation computationnelle et l'Intelligence Artificielle. Ce *modus vivendi* épistémologique est alors d'autant plus acceptable que les terres inconnues qui séparent l'investigation neurologique et le cognitivisme « mentaliste » sont immenses. En 1986, Pylyshyn soulignera à nouveau que les tentatives d'adéquation entre des descriptions en termes psychologiques et en termes physiques sont un échec (« the general failure of perceptual psychology to adequately describe stimuli in physical terms »⁹). Constat suffisamment manifeste pour que les progressions ascendantes et descendantes n'entrent pas de sitôt en éventuel contact.

Toutefois, un observateur aussi informé que le philosophe D. Dennett dessine en ces termes l'avenir de ces recherches :

« En principe, les deux stratégies (ascendante et descendante) sont légitimes puisqu'elles devraient finir par avoir accompli exactement la même tâche. Au bout du compte, les deux parties veulent comprendre la relation entre le cerveau et l'esprit. Vous pouvez commencer par l'esprit et procéder en descendant, soit commencer par des morceaux de cerveau et procéder en remontant. Si vous faites une analogie avec la construction d'un chemin de fer transcontinental, deux équipes commencent à chaque bout et leur plan est de se rencontrer quelque part au milieu »¹⁰.

Cette métaphore rassurante, que l'on pourrait qualifier de « double chantier », suppose que ces deux progressions s'inscrivent *dans un même espace épistémique* bipolaire et que les chercheurs se réfèrent à une commune triangulation de leurs avancées. S'il en était ainsi, une jonction du discours neurologique et du discours cognitif annulerait à terme le trait d'union problématique du corps et de l'esprit, autrement dit supprimerait la frontière entre psychologie et neurologie. Entre la biologie moléculaire et les sciences de l'esprit, un continuum garantirait, niveau après niveau, la naturalisation de la pensée.

9. Cité par Jean PETITOT, « Le Physique, le Morphologique, le Symbolique : remarques sur la vision », *Revue de synthèse*, n^o spécial « Sciences cognitives : quelques aspects problématiques », IV^e S., 1-2, janv.-juin 1990, p. 141.

10. Daniel DENNETT, cf. « States of Mind », ed. Jonatan MILLER, Londres, British Broadcasting Corporation, 1983, p. 63, cité par Pierre JACOB, « Questions pour la table ronde sur l'épistémologie des sciences cognitives », in *Sciences de la Cognition. Grands Colloques de prospective*, Paris, 28-31 janv. 1991, Paris, Ministère de la recherche et de la technologie, 1991, p. 145.

3. — LIMITES DE LA MÉTAPHORE DU « DOUBLE CHANTIER »

Cependant, pour admissible que soit cette représentation de l'état des recherches en 1983, elle n'entraîne pas *ipso facto* la complémentarité théorique de ces démarches cheminant en sens contraire. Refuser avec Requin l'hégémonie d'un réductionnisme ascendant traduit le refus de physicaliser, en partant du cortex, les processus cognitifs. Mais on s'aperçoit vite que le procès fait aux neurosciences de prétendre *réduire* le cognitif au neuronal n'annule pas le procès inverse que le neurobiologiste ne manque pas de faire au cognitiviste qui subordonne le neuronal au computationnel. Rien, en effet, n'impose nécessairement que la démarche ascendante voie ses explications s'exténuier au fur et à mesure qu'elle s'approche des processus cognitifs. Un éliminativisme de principe, posant l'esprit comme un pseudo-objet, semble au contraire une attitude stimulante pour le chercheur des neurosciences qui se donne pour objectif de substituer à terme « l'homme neuronal » à l'*homo psychologicus*, comme l'annonce Changeux en cette même année 1983. « L'homme, écrit-il, n'a dès lors plus rien à faire de "l'Esprit", il lui suffit d'être un homme neuronal »¹¹.

« Homme neuronal », cette ambition revient à *éliminer* dans un avenir plus ou moins proche le bien-fondé de la psychologie cognitive. Mais est-on en droit d'établir une symétrie épistémologique qui autoriserait la démarche descendante à adopter à son tour une thèse éliminativiste ? Le cognitivisme « computo-représentationnel » (c'est-à-dire qui se donne pour modèle un système symbolique régi par des algorithmes), tel qu'on le rencontre chez Fodor ou Pylyshyn, ne conteste pas les contraintes d'une implémentation neuronale. Il postule seulement que la structure du cortex est subordonnée aux fonctions cognitives tout comme le matériel (*hardware*) de l'ordinateur « incarne » la computation symbolique du logiciel (*software*).

En revanche, les thèses éliminativistes (pour qui le pôle mental n'a pas plus de réalité que le phlogistique) se présentent comme le symptôme d'une *non-complémentarité* de principe des démarches ascendante et descendante et du caractère non homogène et provisoire du champ bipolaire cortex-esprit. Les neurosciences, bien loin de reconnaître que leur domaine trouve ses bornes à mi-chemin du parcours qui sépare le cerveau de l'esprit comme le voulait Dennett en 1983, se proposent « d'aménager notre des-

11. Jean-Pierre CHANGEUX, *L'Homme neuronal*, Paris, Fayard, 1983, p. 227.

cription du cognitif en une théorie cohérente qui, au moyen d'un langage formel explicite, pourra alors, et alors seulement, se réduire à une théorie neurale »¹². Leur projet n'est pas de parcourir la moitié de la distance séparant les deux pôles *mais de l'annuler* afin de parvenir à une théorie cognitive réductible à une théorie neurale.

Il faut donc reconsidérer la notion de niveau dont on voit l'ambiguïté si on l'interprète comme une succession ordonnée de structures fonctionnelles qui, du cerveau à la pensée, seraient susceptibles d'être décrites par un même appareil conceptuel au sein d'une conception neuronale globale. Aussi, plutôt que de conserver la métaphore programmatique du « double chantier », convient-il de reconnaître avec Michel Imbert qu'il existe « un va-et-vient constant de l'étude des mécanismes moléculaires et cellulaires du fonctionnement du système nerveux à celle des processus psychologiques les plus subtils » et de retenir « l'idée d'une relative autonomie des niveaux d'explication »¹³.

Ce constat prudent revient à admettre qu'entre neurosciences et sciences cognitives, on cherche à substituer une « réduction transdisciplinaire réciproque ». Ce va-et-vient suppose que tels processus cognitifs ne sont pas « sous-tendus » par telles modifications de l'activité neuronale qui leur seraient « sous-jacentes » ou en constitueraient le « substrat »¹⁴. Ces expressions commodes — si souvent rencontrées — règlent à trop bon compte la question des niveaux par le recours à la métaphore facile d'une architectonique. À cette image où le niveau supérieur (cognitif) reposerait sur des fondations neurologiques, on propose de substituer celle d'une sorte d'horizontalité non contiguë des domaines.

Si le couple cerveau-pensée fait problème pour les sciences cognitives comme pour les neurosciences, c'est précisément en ceci qu'on cherche, de part et d'autre, à *supprimer la frontière* que constitue l'implémentation du « logiciel » dans le « matériel » plutôt qu'à en pérenniser le tracé. Certes, les sciences cognitives tirent leur origine d'un refus du dualisme cartésien mais elles achoppent dans leur projet puisque la « machine de Turing » et la machine corticale sont très loin de superposer en une seule leurs deux images. Comme le souligne en 1983 J. Paillard, entre l'ordinateur et le cerveau, le contraste est saisissant. Il s'agit pourtant de deux machines, l'une formelle, l'autre réelle, donc de *deux systèmes rationnels privés de subjectivité* que l'on cherche à confondre afin d'annuler le trait d'union qui séparait le corps et l'esprit et de désaffecter ces deux termes au profit d'un nouveau. Mais lequel ? Provisoirement, un fantôme.

12. Michel IMBERT, « Neurosciences et sciences cognitives », in D. ANDLER, *op. cit. supra* n. 6, p. 76.

13. *Ibid.*, p. 75.

14. J. REQUIN, *op. cit. supra* n. 2, p. 41.

4. — PÉNÉTRABILITÉ ET IMPÉNÉTRABILITÉ COGNITIVES

Un fantôme dans la machine¹⁵, l'image exprime l'apparition insolite d'un système matériel capable de raisonner avec la même rigueur que l'esprit humain. C'est en 1956 que Herbert Simon et Allen Newell programment sur l'ordinateur Johnniac le *Logic Theorist* qui produit la preuve complète du théorème 2.01 tiré des *Principia Mathematica* de Whitehead et Russell, puis 38 des 52 premiers théorèmes du chapitre 2 des *Principia*. Quelques années plus tard, Newell et Simon^{15bis} insisteront en ces termes sur le choix théorique de l'Intelligence Artificielle :

« Nous ne croyons pas que cette équivalence fonctionnelle entre le cerveau et les ordinateurs signifie qu'il existe une quelconque équivalence à un niveau anatomique plus précis (comme par exemple une équivalence des neurones avec des circuits). Découvrir comment les mécanismes neuronaux réalisent ces fonctions de traitement de l'information dans le cerveau humain est une tâche qui se situe à un autre niveau de la construction théorique. Notre théorie est une théorie des processus d'information impliqués dans la résolution de problèmes, elle n'est pas une théorie des mécanismes neuronaux ou électroniques du traitement de l'information. »

On ne peut plus clairement reconnaître qu'il existe alors une frontière entre les mécanismes neuronaux et les mécanismes cognitifs. Cette démarche est exemplaire de la longue période de près de vingt-cinq ans (1956-1980) au cours de laquelle se multiplient les langages de programmation qui concernent le niveau le plus élevé de la simulation de la pensée par l'ordinateur. Ainsi, les neurones « humides » ne sont pas la condition exclusive d'une conduite intelligente. Il est démontré que le fonctionnement cortical est d'ordre calculatoire (*i.e.* computationnel) et l'on est désormais fondé à parler d'identité entre les *systèmes symboliques physiques* artificiels et naturels, si par « naturels » on entend le cortex humain. Devenu objet de connaissance expérimentalement falsifiable, le fonctionnement cognitif de type logico-mathématique (résolution de problèmes) est « cognitivement pénétrable » mais cette incarnation *logique* de l'esprit dans le cortex ne résout pas la question d'une incarnation *biologique* de la pensée dans le corps. Comme le souligne Fodor en 1983 dans *Modularity of*

15. Jacques BOUVERESSE, « Le fantôme de la machine », in *La Parole malheureuse ?*, Paris, Minuit, 1971.

15 bis. Cf. *Psychological Review*, 65, 1958, p. 151-166.

Mind, « en général, l'interface entre un ordinateur et le monde est un être humain (raison pour laquelle les ordinateurs sont des ordinateurs et non des robots) »¹⁶.

Cognitivement « pénétrable », l'équivalence fonctionnelle du cortex et de l'ordinateur limite la validité de la simulation aux seuls contenus représentationnels fournis à l'ordinateur. Elle ne rend pas compte du procès par lequel le système nerveux capture l'information provenant de l'environnement, c'est-à-dire *fait fonction d'interface*, au niveau de la perception et à celui de l'acquisition des connaissances. C'est en effet autour de la perception que la remise en cause des frontières va prendre tournure. À Herbert Simon qui déclarait encore en 1980 : « Tout ce qui se passe d'intéressant en matière de cognition dépasse le seuil des cent millisecondes — le temps que vous mettez à reconnaître votre mère », Douglas Hofstadter répond en 1985 :

« Je suis du parti diamétralement opposé : tout ce qu'il y a d'intéressant en matière de cognition se passe en deçà du seuil des cent millisecondes — le temps que vous mettez à reconnaître le visage de votre mère. À mon sens, la question maîtresse de l'I.A. tient en ces termes : "qu'est-ce qui peut bien se produire qui vous accorde de convertir cent millions de points rétinien en un mot unique, *mère*, en un dixième de seconde?". C'est à la perception que tout se joue ! »¹⁷.

Et il ajoute : « Il faut sortir du rêve de Boole. »

Une telle contestation symbolise la fin d'une époque durant laquelle la simulation de l'activité logico-mathématique de l'esprit faisait de l'ordinateur un *émule*. L'équivalence fonctionnelle du calculateur et du cortex se limitait aux seules performances dont l'homme avait l'impression d'avoir *consciemment* la maîtrise (par exemple, le jeu d'échecs). La pénétrabilité cognitive des processus constituait jusqu'alors le champ de la simulation et la logique de Boole (qui régit la calculabilité des opérations mathématiques) suffisait à garantir l'équivalence de l'homme rationnel et du logiciel de la machine. Remettre en cause l'hégémonie booléenne comme le fait Hofstadter signifiait que les sciences cognitives étaient invitées à *réformer le statut de l'ordinateur* dans ses modélisations. D'émule fonctionnel qu'il était des opérations cognitives « pénétrables », il allait devenir l'instrument grâce auquel on pourrait simuler les opérations cognitivement « impénétrables » à la conscience, celles qui s'effectuent en moins de 100 millisecondes.

16. Jerry A. FODOR, *The Modularity of Mind*, Cambridge, MA, M.I.T., 1983, trad. fr. *La Modularité de l'esprit*, Paris, Minuit, 1986, p. 60.

17. Douglas HOFSTADTER, « Cognition, subcognition. Sortir du rêve de Boole », *Le Débat*, n^o spécial « Une Nouvelle Science de l'esprit », 47, nov.-déc. 1987, p. 26.

5. — LA « RÉVOLUTION » CONNEXIONNISTE

Un tel programme se présente comme un événement épistémologique qui remet en cause la problématique bipolaire des vingt années précédentes.

Le connexionnisme se défend de vouloir occuper, entre le pôle neuronal et le pôle mental, les vastes régions encore inviolées en vue d'accélérer la jonction future des chantiers ascendant et descendant. Ainsi décrit, le connexionnisme revêtirait l'apparence d'un recours inespéré venant confirmer le bien-fondé de la bipolarité. On a vu ce que la métaphore du « double chantier » avait d'inadéquat. Elle offrait entre autres l'inconvénient de conférer aux neurosciences la charge de rendre compte en termes physicalistes des processus cognitifs « impénétrables » (ceux qui ne sont pas justiciables d'un traitement symbolique) et de réduire ceux-ci aux seules contraintes de l'implémentation.

En postulant que l'activité cognitive se déploie pour l'essentiel *en dessous* du seuil de la conscience (en jargon, le niveau « subdoxastique » ou « subsymbolique »), les connexionnistes refusent l'identité que les cognitivistes « classiques » établissaient entre cognition et représentation symbolique. Cette relation, jugée par eux indispensable pour rendre compte des processus discursifs et inférentiels n'est plus fonctionnelle si l'on se donne pour champ d'étude les mécanismes à l'origine de la perception ou de l'apprentissage. Était du même coup contestée l'attribution aux seules neurosciences du soin de réduire ces comportements à une architecture de systèmes modulaires (précâblés) grâce auxquels le vivant traite les informations lui parvenant des capteurs sensoriels pour aboutir à la perception (des cent millions de récepteurs rétiniens à la reconnaissance du visage maternel).

Ce champ d'étude — jusqu'alors dévolu aux seules neurosciences et à leurs techniques d'investigation — devient le domaine privilégié de la recherche connexionniste dans un langage qui *n'est pas* celui de la biologie moléculaire tout en offrant, comme nous allons le voir, une modélisation plus compatible avec les données de la neurologie (tout en restant compatible avec les inférences booléennes du cognitivisme « classique »). Ainsi, au partage des compétences neurologiques et cognitives fondé sur une hiérarchie de niveaux, allait se substituer la prétention d'explorer dans un nouveau langage le domaine subsymbolique que le cognitivisme classique avait laissé aux neurosciences le soin de décrire.

Les modèles symboliques du cognitivisme, en épousant selon des chaînes linéaires le développement logique des opérations les plus abs-

traies de l'esprit, en épousaient en même temps la rigidité. Performants quand il s'agissait de simuler les cheminements logico-mathématiques, ils n'étaient pas à même de restituer les caractéristiques les plus ordinaires de l'activité mentale.

« Les systèmes de règles butaient sur leur "fragilité", leur manque de flexibilité, leur difficulté, sur l'apprentissage en fonction de l'expérience, sur leur incapacité à généraliser correctement, sur la spécificité des domaines et sur les inefficacités dues à la recherche sérielle dans de grands systèmes »¹⁸.

Ce qui fait le plus clair de la cognition — sa souplesse adaptative, son peu de rigueur déductive, sa capacité de généraliser à partir d'indices — échappait aux contraintes rigoureuses d'un modèle booléen. Ces limitations étaient des obstacles à une pénétration « descendante » de ce formalisme et contribuaient à pérenniser la béance séparant les neurosciences de la psychologie cognitive.

Naturaliser la cognition rendait donc nécessaire, aux yeux des connexionnistes, l'emploi d'un *autre* formalisme, non linéaire celui-ci, qui tiendrait compte à la fois du caractère parallèle des réseaux neuronaux et de la capacité adaptative des processus cognitifs, notamment dans le comportement perceptif et dans l'apprentissage de règles (par exemple, comment on apprend une langue). Convaincus que la rigidité du formalisme booléen lui interdirait d'avancer dans ce champ d'étude, les connexionnistes seront frappés par l'analogie de ces contraintes souples avec celles qu'on rencontre en physique dans des systèmes dynamiques tels que les verres de spin (Hopfield, 1982). Là où le cognitivisme computationnel était obligé d'implémenter des règles, les connexionnistes supposent que celles-ci, calculées *par* le système, sont assimilables à des états d'équilibre locaux.

« L'idée fondamentale, résumée W. Bechtel et A. Abrahamsen, est qu'il existe un réseau d'unités élémentaires ou nœuds dont chacun a un certain niveau d'activation. Ces unités sont connectées entre elles, de sorte que les unités actives excitent ou inhibent les autres unités. Le réseau forme ainsi un système dynamique qui, lorsqu'on lui fournit une entrée initiale, propage les excitations parmi ses unités. Pour comprendre comment un système cognitif accomplit une tâche cognitive, il faut fournir une interprétation. On le fait généralement en considérant les activations initiales fournies au système comme spécifiant un problème, et la configuration stable produite à la fin du traitement comme la solution que le système apporte au problème »¹⁹.

Ni modélisations du système nerveux en tant que tel, ni modélisations de l'activité symbolique, les formalisations connexionnistes « semblent être

18. W. BECHTEL, A. ABRAHAMSEN, *op. cit. supra* n. 1, trad. fr. p. 26.

19. *Ibid.*, p. 10.

inspirées de plus en plus par des considérations mathématiques et de moins en moins par des considérations neuronales »²⁰. Avec eux, le « contraste saisissant » du cortex comparé à l'ordinateur perd de sa pertinence tout comme en perd l'opposition polaire du cerveau et de l'esprit.

« Dans la machine virtuelle symbolique, remarque Schyns, les processus cognitifs sont conçus comme des manipulations de structures symboliques qui désignent des objets du monde extérieur. La machine virtuelle connexionniste conçoit la cognition comme l'émergence d'une micro-activité dominée par les concepts de minimisation d'énergie et de satisfaction de contraintes »²¹.

Ce faisant, les modélisations connexionnistes prétendent apporter des réponses qui ne soient pas incompatibles avec les données neurologiques *tout en restant computationnelles* comme le sont les modèles de traitement symbolique.

L'intérêt est alors de savoir non pas *où* mais *comment* situer le connexionnisme par rapport aux neurosciences et à la psychologie cognitive. Si, en effet, il conteste la métaphore du « double chantier » de Dennett en raison de la spécificité de son nouveau formalisme mathématique (qui ne dépend plus de la notion de représentation symbolique), et si, d'autre part, le formalisme physicaliste des neurosciences n'est pas réductible au formalisme du cognitivisme classique, il faut bien reconnaître que ces trois démarches *ne se situent pas* dans un espace théorique homogène.

6. — NIVEAUX D'ANALYSE ET NIVEAUX D'EXPLICATION

On se rappelle que Requin avait mis en doute la pertinence des « niveaux d'organisation » définis hiérarchiquement en fonction du « grain » de l'analyse choisi. Il insistait sur le fait que la constitution de ces niveaux ne relevait pas d'une observation empirique liée aux seules capacités techniques d'investigation mais que celles-ci dépendaient d'une modélisation dont elles éprouvaient la validité. C'est ainsi que l'objet des neurosciences peut être décrit du niveau le plus élémentaire au niveau le plus molaire (molécules, synapses, neurones, réseaux locaux, couches et colonnes, cartes topographiques et systèmes). Mais ces observables neuronaux laissent ouverte la possibilité de les décrire comme processus non physica-

20. Selon Paul SMOLENSKY en 1988, in *ibid.*, p. 304.

21. Philippe SCHYNS, « Psychologie de synthèse : les métaphores de l'esprit calculateur », in *Intelligence naturelle et Intelligence artificielle*, dir. Jean-François LE NY, Paris, Presses universitaires de France, 1993, p. 357.

listes, inobservables, sous réserve que la machine virtuelle choisie pour les simuler en respecte au mieux la structure et le fonctionnement. On sait que les modélisations de traitement symbolique en étaient incapables et abandonnaient l'implémentation des processus cognitifs aux neurosciences. Les modèles connexionnistes vont apporter la preuve que l'implémentation peut relever *elle aussi* d'un discours qui se distingue de celui des neurosciences. Ce faisant, ils remettent en cause la métaphore conciliatrice du « double chantier » et l'idée qu'à mi-chemin des parcours ascendant et descendant l'avenir des recherches établirait une (énigmatique) jonction.

Le connexionnisme se pose d'emblée comme une nouvelle approche théorique du problème corps-esprit en contestant que l'implémentation soit du seul ressort des neurosciences. Du même coup, il remet en cause l'idée d'une frontière séparant neurosciences (le *hardware* de l'implémentation) et psychologie cognitive (le *software* computo-représentationnel).

Pour faire comprendre leur ambition, les connexionnistes font allusion aux travaux de David Marr (*Vision*, Cambridge, MA, 1982) qui distingue trois niveaux d'*explication* (et non plus d'analyse) pour rendre compte de la vision comme traitement de l'information : un *niveau computationnel* qui est celui de la fonction (au sens mathématique) à calculer ; un *niveau de l'algorithme* particulier choisi pour calculer la fonction (qui inclut le choix d'un symbolisme particulier) ; enfin, le *niveau de l'implémentation*. Pour Marr, la raison d'être de ce dernier niveau est d'expliquer comment les représentations (niveau computationnel) et l'algorithme spécifié (niveau de l'algorithme) sont implémentés dans les états et les processus d'un dispositif « sous-jacent » (neurones, puces en silicone, etc.). Dans l'exemple de la vision, Marr propose que chaque niveau d'analyse (des récepteurs rétiniens aux régions corticales, en passant par des systèmes intermédiaires) soit caractérisé par une construction, à partir d'une esquisse primaire pleine (en 2D) au niveau rétinien, puis d'une seconde qui définirait un système coordonné centré sur l'individu qui voit (en 2 1/2D), pour aboutir enfin aux formes et à leur organisation spatiale dans un système coordonné tridimensionnel, centré sur l'objet (en 3D). Chacun des paliers de cette construction représente les états et processus d'une *machine virtuelle*. Ces états et processus *ne sont donc pas définis physiquement* mais comme des entrées et des sorties de processus computationnels calculés à partir d'un plus bas niveau d'analyse. Bref, les trois « niveaux d'explication » de Marr s'appliquent à chacun des « niveaux d'analyse » puisqu'à chacun de ceux-ci la computation, l'algorithme et l'*implémentation* sont requis.

7. — MODÈLES CONNEXIONNISTES ET MODÈLES DE TRAITEMENT SYMBOLIQUES

En évoquant la théorie de Marr, les connexionnistes veulent faire entendre aux neuroscientifiques et aux cognitivistes qu'ils ne proposent pas une modélisation intermédiaire et conciliatrice mais une nouvelle formalisation (mathématique) des processus cognitifs. Mais on pouvait s'attendre à ce que les cognitivistes « classiques », notamment Fodor et Pylyshyn²², contestent le bien-fondé de ce modèle.

Jay G. Rueckl leur fait tenir ce genre de propos :

« Les modèles connexionnistes sont simplement inadéquats en tant que théorie d'un niveau algorithmique. Nous ne pouvons imaginer comment ils pourraient faire X (où X est généralement un problème impliquant un raisonnement abstrait ou l'utilisation du langage). La seule façon de faire X est de manipuler des symboles. Bien sûr, nous sommes d'accord sur le fait que les modèles connexionnistes ont de nombreuses propriétés intéressantes mais celles-ci ne sont importantes qu'au niveau implémentatif. Peut-être le connexionnisme fournira-t-il des explications, du niveau implémentatif au niveau symbolique. Mais s'il en est ainsi, quoi de plus ? La psychologie est dans l'algorithme, pas dans son implémentation »²³.

À quoi, ajoute Rueckl, un connexionniste pourrait rétorquer :

« Bien que les réseaux connexionnistes ressemblent beaucoup aux systèmes neuraux, ils se situent en réalité à un niveau plus abstrait. Bien sûr, ce qu'ils offrent est une conception nouvelle des algorithmes utilisés pour réaliser des processus computationnels — conception qui ne repose pas sur des entités telles que symboles, règles, etc. mais emploie plutôt des concepts comme flux d'activation, force de poids et représentation distribuée. Ainsi, si l'on accepte des modèles connexionnistes comme théories psychologiques valides, ce qu'on fait c'est accepter de nouvelles explications situées au niveau algorithmique. »

Comme on le voit, la question des frontières entre neurosciences et psychologie cognitive, loin d'être réglée, trouve avec l'apparition du connexionnisme un rebondissement qui a l'intérêt d'écarter la lecture du

22. Jerry A. FODOR, Zénon PYLYSHYN, « Connectionism and Cognitive Architecture : A Critical Analysis », *Cognition*, 28, 1988, p. 3-71.

23. Jay G. RUECKL, « Connectionism and the Notion of Levels », in *Connectionism and the Philosophy of Science*, éd. T. HORGAN, J. TENSON, Dordrecht, Kluwer Academic Press, 1991, p. 75.

problème en termes d'exclusion polaire (neural vs mental). Dans ces nouveaux modèles, les représentations symboliques *ne sont plus*, vis-à-vis des implémentations neurales, dans le même rapport que celui du logiciel et du matériel de l'ordinateur. Si ce dernier demeure le médium calculatoire indispensable à la recherche, sa structure n'est plus supposée analogue à celle du cortex (discontinuité entre logiciel et matériel analogue à la discontinuité neural-symbolique). Il devient l'instrument calculateur au service d'une théorie mathématique.

La simulation par machine virtuelle des processus cognitifs, si elle concerne toujours une implémentation dans un système physique (neurone ou puce au silicium) n'est pas nécessairement contrainte à appliquer au système physique naturel du « cerveau humide » (le cortex) la relation rencontrée dans un calculateur d'« architecture von Neumann » entre logiciel et matériel. Tandis que le cerveau est l'effet d'une évolution biologique où structure et fonction sont par nature indissociables, il n'en est pas de même de l'ordinateur dans lequel la structure est un artefact subordonné à la fonction : la machine de Turing matérialisée possède une implémentation contingente, exclusivement justifiée par sa capacité de calculer tout nombre calculable. D'où une double interprétation polémique des modèles connexionnistes : ou bien on limite leur prétention à donner une explication potentielle de l'implémentation dans le « matériel neural », ou bien on les perçoit comme des *compétiteurs* des modèles de traitement symbolique.

Ces deux interprétations ne sont pas le fait des chercheurs en neurosciences qui verraient plutôt favorablement la compatibilité des structures neuronales avec les réseaux parallèles et non linéaires des modèles connexionnistes. Elles émanent des cognitivistes « classiques », irréductiblement attachés à la nature représentationnelle (symbolique) des processus cognitifs, pour lesquels le niveau computationnel « absorbe » entièrement l'emploi des algorithmes. En faisant de ceux-ci un niveau explicatif spécifique pour expliquer les processus subsymboliques, les connexionnistes contestent la frontière que les cognitivistes orthodoxes situaient *en deça* de l'implémentation dans le système physique neuronal. Restait alors aux cognitivistes à faire basculer le connexionnisme par-dessus la frontière et de réduire leur utilité à procurer à l'implémentation une modélisation ingénieuse.

Compétiteur ou simple adjuvant des neurosciences, le connexionnisme est, dans les deux cas, rejeté. Compétiteur du cognitivisme « computationnel », il est accusé de proposer une véritable révolution théorique (un changement de paradigme). Simple adjuvant des neurosciences, ses explications se situent au mauvais niveau (« at the wrong level »²⁴)

24. *Ibid.*, p. 74.

parce que négligeant le caractère essentiellement symbolique des processus psychologiques.

Mais il conviendrait de savoir si les représentations symboliques, donc discrètes, sont la *seule* forme que peut prendre la computation. Si l'on prend l'exemple de la perception, un connexionniste pourra objecter que la reconnaissance du visage de votre mère n'est une forme complètement stabilisée que si elle est l'effet d'une computation à *tous* les niveaux d'analyse (2D, 2 1/2D, 3D dans le modèle de Marr), y compris lorsque votre cerveau produit le mot « mère ».

Les modèles connexionnistes peuvent en outre invoquer leur supériorité explicative dans le domaine de l'apprentissage où les modèles de traitement symbolique sont singulièrement fragiles. Bien plus, certains d'entre eux sont à même d'implémenter des processus de manipulation symbolique. Ainsi, dans les systèmes de production de Tourestsky et Hinton (1986), ce qui peut être décrit en termes de représentation distribuée à un certain niveau d'analyse peut être décrit en termes de symboles et de règles à un autre niveau. Ces capacités explicatives confirment que le statut des modèles connexionnistes résiste à une description bi-polaire des relations entre le neuronal et le cognitif où ils rempliraient une fonction d'interface au seul niveau de l'implémentation.

8. — LE PROBLÈME DES FRONTIÈRES ET LA LIMITATION DES FORMALISMES

Il a été beaucoup question de niveaux dans l'évocation des travaux consacrés depuis une trentaine d'années au problème corps-esprit. On a pu constater que depuis l'apparition des modèles connexionnistes, le partage des descriptions entre neurosciences et psychologie cognitive ne pouvait se ramener à une approche complémentaire de la structure neuronale et de la fonction cognitive. À la métaphore du « double chantier » qui avait cours du temps où la modélisation en termes symboliques dominait les sciences de l'esprit, le développement des modèles connexionnistes oppose un nouveau formalisme dans lequel l'implémentation *ne peut plus servir de frontière* entre neurosciences et psychologie cognitive. En déplaçant vers les processus subsymboliques la problématique cognitive, le connexionnisme vise à mettre un terme à l'analogie peu parcimonieuse de l'ordinateur et du cortex dans laquelle le neurone est à la pensée ce que le matériel est au logiciel.

« Pour résumer l'opposition entre les approches symboliques et subsymboliques, déclare Smolensky au Colloque de Cerisy en 1987, nous dirons d'abord

que dans le paradigme subsymbolique, on a des lois fondamentales qui sont des équations différentielles et non des procédures de manipulation de symboles; que les systèmes décrits sont des systèmes dynamiques, et non des machines de von Neumann; et enfin que la catégorie mathématique dans laquelle se développent ces formalismes est la catégorie du continu et non celle du discret : une mathématique différente entre en jeu »²⁵.

« Les sciences cognitives de l'an 2000 seront des sciences cognitives très différentes de celles qui auraient émergé en l'absence du néo-connexionnisme »²⁶, mais ce dernier ne prétend pas se substituer aux modélisations de traitement symbolique dont on a comparé le pouvoir explicatif à celui du système newtonien. L'apparition du système quantique, en apportant des garanties à la fois théoriques et empiriques que le système newtonien ne soupçonnait pas, n'a pas annulé pour autant la validité de celui-ci. Il en irait de même entre les modèles de traitement symbolique et le connexionnisme²⁷.

S'il en est ainsi, la question de la frontière entre neurosciences et psychologie cognitive ne s'est pas simplement déplacée : elle ne distingue plus des domaines sur lesquels s'appliqueraient des formalismes mais des formalismes auxquels des domaines doivent leur existence.

Alain TÊTE,
Université de Provence,
CREPCO, URA C.N.R.S. 182,
29, av. Robert-Schuman,
13621 Aix-en-Provence Cedex 1
(juin 1994).

25. P. SMOLENSKY, « I.A. connexionniste, I.A. symbolique et cerveau », 1987, in D. ANDLER, *op. cit. supra* n. 6, p. 88.

26. W. BECHTEL, A. ABRAHAMSEN, *op. cit. supra* n. 1, p. 29.

27. P. SMOLENSKY, *art. cit. supra* n. 25, p. 95.