



**HAL**  
open science

# La psychologie cognitive survivra-t-elle aux sciences cognitives ?

Guy Tiberghien

► **To cite this version:**

Guy Tiberghien. La psychologie cognitive survivra-t-elle aux sciences cognitives?. Psychologie Française, 1999, 44, pp.265-283. hal-00089284

**HAL Id: hal-00089284**

**<https://hal.science/hal-00089284>**

Submitted on 16 Aug 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Publié dans

Tiberghien, G. (1999). La psychologie cognitive survivra-t-elle aux sciences cognitives ? In J-F. Richard, & G. Tiberghien, (Eds.). *Epistémologie et Psychologie. Psychologie Française*, 44 (3), 265-283.

## **La psychologie cognitive survivra-t-elle aux sciences cognitives ?**

**Guy Tiberghien**  
**Institut des Sciences Cognitives, Lyon**

- Bien que les sciences cognitives aient atteint en quelques décennies une maturité scientifique et une reconnaissance institutionnelle difficilement contestables, leur définition est toujours problématique et leur origine même est encore discutée. Cette situation n'est paradoxale qu'en apparence étant donné la déstabilisation épistémologique considérable engendrée par cette nouvelle approche dans la rassurante classification du savoir héritée d'un positivisme dominateur.

- C'est tout d'abord la définition même des sciences cognitives qui pose problème. Cela va du simple déni de cognition à une réduction plus ou moins subtile du concept de cognition en passant par un oecuménisme de bon aloi. On peut, par exemple, affirmer que les sciences cognitives ont toujours existé et que cette expression est purement "tautologique" . Certains plaident pour une définition consensuelle des sciences cognitives, vaste "archipel" de disciplines (psychologie, informatique, neurobiologie, linguistique, etc.) dont la coopération devrait permettre d'élucider les processus de la connaissance (Vergnaud, 1991). Cette définition, en évitant tous les problèmes, est sans doute recommandable pour les institutions qui ont à gérer les sciences cognitives, mais elle est de peu d'utilité scientifique. Il y faut une prise de risque plus grande et un point de vue inévitablement plus réducteur. En intelligence artificielle on aime, par exemple, à parler de "système" cognitif, voire "d'être intelligent" ou d'intelligence "en général" (Simon & Kaplan, 1989) ; les philosophes de la cognition préfèrent étudier "l'esprit" ou "l'esprit humain" (Andler, 1992). Cet esprit là peut être étudié sans référence explicite au cerveau, comme le font classiquement les philosophes et les psychologues, mais il peut aussi être mis en relation avec lui, comme en neuropsychologie et, plus généralement, dans les neurosciences. Ces dernières prises de position sont plus engagées, ce qui est peut-être un gage de progrès, mais elles sont toujours exposées à quelque forme de réductionnisme théorique ou disciplinaire.

- Ce débat sur la définition des sciences cognitives n'est pas indépendant, bien sûr, de la question des origines des sciences cognitives. La façon dont on fait l'histoire n'est pas une question mineure. Elle détermine, au contraire, la position relative des diverses disciplines au sein des sciences cognitives, le type de définition qui en résulte et les problèmes empiriques et théoriques prioritaires. Ainsi, il n'est pas indifférent de situer l'origine des sciences cognitives au moment de l'émergence de la "première cybernétique" — après la seconde guerre mondiale — ou de la situer lors du débat entre gestaltistes, behavioristes et néo-behavioristes — entre la première et la seconde guerre mondiale. Dans la premier cas,

le rôle et la place de l'informatique sont décisifs et celui de la psychologie relativement éclipsé (Dupuy, 1994, p. 27). Dans le second cas, la source des sciences cognitives est inséparable du développement de la psychologie cognitive dont le rôle central est alors évident. Le cognitivisme est, en effet, une théorie psychologique bien antérieure aux sciences cognitives ; il en a d'ailleurs largement préparé l'avènement en rompant progressivement avec le behaviorisme entre 1920 et 1940 (Gardner, 1985/1993). A fortiori, il ne saurait être évidemment incarné dans une théorie unique ou dans les travaux d'un seul auteur.

- En réalité, à l'origine, il n'y a pas une mais deux révolutions cognitives: la première, dans laquelle la psychologie a joué un rôle majeur, voit l'émergence d'un nouvel objet scientifique, la représentation mentale ; la seconde, inséparable de l'essor de l'informatique, pose que la représentation mentale peut-être décrite par un langage formel et simulé par un programme informatique mémorisé. La première révolution est théoriquement fondamentale et indissociable de l'histoire de la psychologie scientifique. La seconde révolution est quasi-méthodologique, plus contingente, si l'on préfère, car elle a été provoquée par l'évolution "technologique" de nos artefacts de modélisation et de simulation.

- Il est donc difficile de nier, bien qu'on puisse la taire, une contribution historique majeure de la psychologie dans le développement des sciences cognitives. Cependant l'éclatant succès des sciences cognitives soulève de nouvelles interrogations. La psychologie cognitive a-t-elle encore un avenir ou sera-t-elle complètement réduite, à terme, aux sciences cognitives ... sera-t-elle même "dévorée", pour reprendre l'expression de Gardner, par les neurosciences cognitives? Les craintes des psychologues de la cognition répondent souvent à certaines déclarations réductionnistes, les premières n'étant pas toujours justifiées et les secondes parfois prématurées (Gardner, 1985/1993, pp. 326-328 ; Tiberghien & Jeannerod, 1995, p. 181). L'évolution récente montre, au contraire, une extension considérable des méthodes, des paradigmes et des modèles psychologiques à d'autres champs des sciences cognitives. Ce n'est pas, semble-t-il, la psychologie cognitive qui est réduite à une description neurophysiologique. Au contraire, les neurosciences ne peuvent plus faire aujourd'hui l'économie d'une approche cognitive, la seule qui puisse, à terme, rendre intelligible son objet d'étude, le cerveau. Quels que soient les aléas institutionnels, la psychologie cognitive ne disparaîtra pas — scientifiquement, j'entends — car elle a été à l'origine même des sciences cognitives, son développement en est indissociable et elle est maintenant au coeur même des neurosciences cognitives. D'ailleurs, c'est peut-être la psychologie cognitive qui est en train d'assimiler progressivement les formalismes computationnels et la connaissance du fonctionnement cérébral, ce qui lui permet d'interagir, de façon de plus en plus pertinente, avec les autres disciplines dans une science de la cognition qui deviendra sans doute de plus en plus intégrée.

### **De la psychologie philosophique à la psychologie cognitive**

- C'est un fait bien connu que la psychologie scientifique ait eu à rompre avec la philosophie afin de se réaliser pleinement. Cette rupture commença à la fin du dix-neuvième siècle mais elle ne fut vraiment accomplie, en France, qu'au milieu de ce siècle. Le libre développement de la psychologie l'a donc conduit à rejeter radicalement l'antipsychologisme kantien. Ce rejet l'a conduit à adhérer à un empirisme plutôt radical et à manifester, de façon durable, une relative méfiance à l'égard de toute réflexion philosophique. Cette philosophie "spontanée" des psychologues, préférant le phénomène à l'essence, s'illustra de façon exemplaire dans le programme behavioriste. S'il était impossible de falsifier les idées pures, il restait possible d'expérimenter sur les comportements.

Toutefois, plusieurs philosophes (Brentano, Husserl, Von Ehrenfels) n'adhéraient pas au pessimisme psychologique de Kant. Cela laissait déjà augurer de nouvelles rencontres avec la psychologie scientifique et l'influence de la gestalttheorie en portera d'ailleurs témoignage. Plusieurs décennies plus tard, les sciences cognitives vont même contraindre la psychologie scientifique à se confronter, une nouvelle fois, à la philosophie sur des questions aussi fondamentales que l'intentionnalité ou la conscience (Pacherie, 1993). Nous reviendrons sur cet étonnant retour des choses qui peut apparaître, au choix, soit comme une «nouvelle» alliance épistémique soit comme une «revanche» du rationalisme philosophique sur l'empirisme psychologique.

- A la fin du siècle dernier, la réflexion rationnelle, de nature introspective, cessa d'être la méthode d'exploration privilégiée de la psychologie. Il lui fallait un substitut, ce fut la méthode expérimentale. La psychologie scientifique naissante a littéralement été subjuguée par les succès des sciences physiques et naturelles (Donders, Helmholtz) et les avancées des mathématiques (Boole, Galton). Tout semblait possible et l'expérimentation, alliée aux méthodes statistiques, allait permettre d'édifier une psychophysique objective (Fechner, Weber), une chronométrie mentale (Donders), et les bases de ce qui allait apparaître comme une nouvelle "chimie mentale" — pour reprendre une expression qui allait connaître son succès bien des années plus tard (Estes, 1960). La psychologie allait devenir physiologique, fonctionnaliste et associationniste . Tout était prêt pour l'irrésistible ascension du behaviorisme.

L'hégémonie de la psychologie du comportement allait être assurée pendant la première partie de ce siècle. Le behaviorisme allait imposer un concept clé, le comportement, un paradigme opérationnel, le schéma S-R, et un programme de recherche, l'apprentissage. Mais que l'on ne se méprenne pas, la mise entre parenthèses méthodologique des réalités mentales constituait une des conditions nécessaires à leur étude expérimentale ultérieure. Cette "révolution copernicienne" du behaviorisme, selon la belle expression de Goussard a consisté à dessaisir, par une sorte de restriction expérimentale, les centres au profit de la périphérie (1959, p. 117). Il en a coûté une nécessaire rupture avec la psychologie philosophique. Mais la pierre de touche de cette rupture était bien la possibilité d'un contrôle expérimental des phénomènes psychologiques et non la mise à l'encan définitive des phénomènes mentaux, de la conscience ou de l'intentionnalité.

- Le behaviorisme progressa alors rapidement dans la compréhension des phénomènes de conditionnement et d'apprentissage, animal et humain. L'apogée de cette entreprise peut être précisément datée en 1943, année de la publication par Hull de son "Principles of Behavior". Cette théorie quasi-formalisée de l'apprentissage représente sans doute le point ultime de l'avancée de la psychologie du comportement. Elle ne sera pas dépassée et échouera quand elle sera appliquée aux apprentissages les plus complexes, comme le langage ou les habilités cognitives. Ce sont ainsi les contradictions et les limitations mêmes du behaviorisme qui ont donné naissance au cognitivisme en psychologie.

- La crise du behaviorisme a cependant été largement préparée par des théories dissidentes apparues en psychologie entre 1920 et 1940: la théorie de la Forme (Köhler, Wertheimer), la théorie topologique du champ (Cartwright, Lewin) et l'épistémologie génétique (Piaget). Peu à peu ces théories imposeront un nombre limité d'idées qui contribueront à façonner le concept de cognition: 1) ce n'est pas l'association qui détermine l'organisation du psychisme mais l'inverse ; 2) l'activité psychologique peut être décrite en termes d'attraction-répulsion dans des champs de forces perceptives, conceptuelles et motrices ; 3) le développement de la pensée humaine ne résulte pas d'un processus purement cumulatif d'acquisitions

élémentaires mais obéit à une logique sous-jacente.

Certes, ces attaques extérieures ont peu à peu ébranlé tout l'édifice behavioriste. Mais c'est de l'intérieur même que le behaviorisme a littéralement implosé. C'est tout d'abord Tolman (1925b, 1948) qui a été contraint, l'un des premiers, de postuler des états représentationnels ("cartes mentales") et même intentionnels (réponses vicariantes) pour expliquer les apprentissages latents et l'orientation spatiale chez l'animal. Les hypothèses tolmaniennes ont ainsi engendré un néo-behaviorisme qui s'est de plus en plus émancipé de la règle behavioriste: Hull est amené à postuler des réponses implicites anticipatrices de but entre les stimulus et les réponses (1952) ; les théories médiationnelles réintroduisent une certaine forme d'état représentationnel ou émotionnel, observable "en principe" ou "potentiellement", entre les états objectivement observés de l'environnement et le comportement (Osgood, 1960 ; Spence, 1956).

Tout était donc prêt pour l'avènement de la psychologie cognitive. En 1948, lors de la Conférence Hixon au Caltech, Lashley récuse avec force la psychologie du comportement, impuissante à rendre compte de l'organisation de nos conduites perceptivo-motrices les plus complexes, démunie pour expliquer le langage et acculée à s'auto-détruire en reconnaissant l'importance de variables intermédiaires ou hypothétiques que presque plus rien ne séparait des représentations mentales. Le paradigme cognitiviste pouvait naître et la psychologie cognitive allait rapidement supplanter le behaviorisme et imposer un nouveau concept clé, une nouvelle méthodologie et un nouveau paradigme. C'est cette première révolution cognitive qui va rendre possible l'émergence des sciences cognitives.

### **De la psychologie cognitive aux sciences cognitives**

La psychologie expérimentale, en rupture de ban avec le behaviorisme, a donc contribué à la genèse des rencontres fondatrices de Hixon et de Macy. Dans le creuset de la cybernétique, la psychologie est entrée en contact direct, et frontal, avec la neurophysiologie, la théorie de l'information et l'informatique. De cette rencontre allait émerger la psychologie cognitive dont l'essor deviendra irrésistible dans la décennie 1950-1960. C'est incontestablement d'une révolution dont il s'agit mais elle s'effectuera en deux temps qu'il convient de ne pas confondre.

La première révolution cognitive consistera à poser que l'objet d'étude de la psychologie est la représentation mentale et que le comportement n'est qu'un simple moyen d'accès objectif permettant la reconstruction des propriétés des états mentaux. C'est évidemment l'acte de décès du behaviorisme. La psychologie cognitive admet, de plus, que le psychisme peut être considéré comme un système de traitement de l'information composé de modules fonctionnels autonomes, spécialisés et agencés dans une architecture contrôlée par un système de supervision. C'est l'acceptation d'un modèle de la pensée fondé sur des principes d'organisation séquentielle ou parallèle (processus "bottom-up") et de rétroaction (feedback, processus "top-down") qui consacre l'influence de la théorie de l'information mais aussi celle du projet de la première cybernétique. Ce nouveau paradigme rendra possible l'étude expérimentale de toutes les questions dont l'approche avait été suspendue, ou à peine esquissée, par la psychologie du comportement : la représentation des connaissances et la catégorisation (Anderson, Rosch), l'attention (Broadbent), le langage (Chomsky, Miller), la mémoire (Miller, Quillian) et le raisonnement (Bruner). Les deux décennies suivantes confirmeront le statut dominant du cognitivisme dans la psychologie expérimentale ( pour une revue : Anderson, 1984 ; Pylyshyn, 1982 ; White, 1983) .

La première révolution cognitive est donc inséparable de l'évolution de la psychologie scientifique. Le dire ne conduit pas, bien sûr, à sous-estimer le rôle majeur d'autres disciplines convergentes. Cette révolution est ainsi fortement corrélée avec les progrès de la neurophysiologie et de la neurologie dont les découvertes permettent de poser en des termes nouveaux la question des rapports entre le cerveau et l'esprit. Le bouleversement technologique provoqué par l'apparition des premiers ordinateurs qui réalisent, avec le théorème de Gödel et la machine de Turing, une intégration opérationnelle des mathématiques, de la logique et de la programmation binaire en est aussi un facteur essentiel. Les Conférences de Macy (1946-1953) et Hixon (1948-1960) sont un bon indicateur social et intellectuel de cette première révolution cognitive (pour une revue : Dupuy, 1985, 1994 ; Pélissier & Tête, 1995). Cependant elles n'en sont pas, à proprement parler, le point d'origine. Elles sont plutôt le révélateur et l'amplificateur de la crise interne du behaviorisme et elles demeureraient peu compréhensibles si elles n'étaient pas mises en relation précise avec l'histoire de la psychologie scientifique et, en particulier, avec l'antagonisme théorique, beaucoup plus ancien, entre behaviorisme et gestalttheorie. On ne devrait donc pas être surpris, outre mesure, de constater que, dès les années 20, les concepts de cognition, d'intention, de but, et même de conscience, étaient déjà au centre des préoccupations de certains behavioristes . Ce sont précisément ces concepts et ces thèmes qui vont être l'objet de controverses permanentes non seulement entre behavioristes mais aussi avec les gestaltistes. Ce sont eux aussi qui vont constituer "l'amorce" de cette première révolution cognitive et alimenter largement les premiers débats fondateurs. D'ailleurs, à l'origine, la cybernétique de Rosenbluth, Wiener, et Bigelow est encore fortement marquée par le behaviorisme psychologique et McCulloch, lui-même, se donne bien un projet psychobiologique d'inscription de l'esprit dans les neurones. Mais la conquête de la psychologie par la cybernétique échoua. Dans l'ensemble, la psychologie scientifique, refusa la forme de néo-behaviorisme neuronal que lui suggérait la première cybernétique et adhéra au cognitivisme.

- Si la psychologie cognitive et la cybernétique résultent de cette première révolution, les sciences cognitives -- au sens où nous les concevons aujourd'hui -- étaient encore dans les limbes. Il faudra attendre le milieu des années 50 pour que les intenses débats sur les relations "cerveau-esprit-machine" aboutissent à la prise de conscience qu'une nouvelle science était peut-être en marche. Bien que le premier emploi de l'expression "sciences cognitives" puisse être précisément daté (1970), c'est en 1956 que le programme scientifique des sciences cognitives émerge lors d'une importante Conférence au MIT, à Cambridge (MA). Participent à cette Conférence, entre autres, Chomsky, Hubel et Wiesel, Miller, Newell & Simon,. Ce sont incontestablement la date et le lieu emblématiques de la seconde révolution cognitive .

- En effet, la première révolution cognitive résulte d'une double rupture, d'une part avec le behaviorisme psychologique, et d'autre part avec le néo-behaviorisme de la première cybernétique. Mais s'il devenait possible d'étudier expérimentalement la représentation mentale, il fallait nécessairement en proposer une description opérationnelle. C'est ce que réalisa la seconde révolution cognitive en décrivant les représentations mentales sous la forme de symboles dont l'inscription physique dans le cerveau était postulée. L'esprit pouvait alors être présenté comme le produit d'une manipulation formelle des symboles opérée par le cerveau, assimilé lui-même à un système de traitement de l'information, à une machine computationnelle. La pensée était donc structurée comme un langage formel ... la pensée était elle-même un langage, un "mentalais" (Fodor & Pylyshyn, 1988). Si le comportement ne permettait pas d'atteindre la pensée, il fallait bien inventer l'outil épistémique permettant de l'objectiver. En réduisant la pensée à un langage, le cognitivisme faisait coup double: il définissait un véritable "comportement" logique et retrouvait ainsi un indicateur observable

et pouvant être simulé sur un ordinateur.

Sous l'influence de la première cybernétique, la science cognitive aurait ainsi pu être d'emblée connexionniste et neuronale, en continuité avec le néo-behaviorisme, mais cette option théorique fut provisoirement abandonnée. La représentation mentale, symbolique, devint le concept unificateur des sciences cognitives, voire son dogme incontestable. Les sciences cognitives originelles, ou "de stricte observance", ne peuvent plus dès lors être dissociées de la psychologie cognitive dont l'évolution va être également, et pour un temps, très influencée par le développement de l'Intelligence Artificielle (IA) . Les sciences cognitives "représentationnelles" vont être dominantes pendant près d'une vingtaine d'années, jusqu'au milieu des années soixante-dix. Mais elles vont être contestées, à partir de cette date, par le programme connexionniste qui réalise une sorte de revanche posthume de la première cybernétique sur le cognitivisme orthodoxe, une révolution dans la révolution.

Le connexionnisme peut être considéré, sous un certain angle, comme un retour aux origines néo-behavioristes et cybernétiques des sciences cognitives. Son essor s'explique d'abord par les difficultés rencontrées par le cognitivisme représentationnel et, en particulier, par sa difficulté à modéliser les processus de contrôle des architectures modulaires de la cognition. L'essor de la psychologie mathématique a également joué un rôle déterminant dans l'apparition du programme connexionniste . Il est ainsi beaucoup plus proche des idées de McCulloch et de Lashley, que de celles de Von Neumann et de Simon. La cognition n'est plus décrite en termes de représentations symboliques mais en termes d'états globaux d'un réseau d'unités de traitement en interconnexion. Ce sont des règles de changement local qui déterminent le changement d'état global du réseau et ses propriétés émergentes s'adaptent progressivement aux variations de l'environnement. Les symboles du programme "représentationnel" étaient manipulés par des computations logiques, la computation des états locaux d'un réseau de neurones est purement numérique. En d'autres termes, c'est l'ordinateur digital qui est la métaphore de référence des sciences cognitives orthodoxes, alors que c'est le cerveau qui est la métaphore de référence du connexionnisme (Hinton, 1989 ; McClelland, 1986). Au fond, le connexionnisme est un véritable behaviorisme neuronal en ce sens qu'il "remplit" la boîte noire d'un réseau de connexions dont les états sont, en principe, des observables .

Sur le plan épistémologique, l'opposition entre le "représentationnalisme" et le connexionnisme semble totale. Toutefois au niveau des programmes de recherche, les différences paraissent beaucoup moins contrastées. En effet, le problème de l'organisation et de l'émergence des représentations symboliques (référence, intentionnalité), à partir des propriétés sub-symboliques d'un réseau, demeure entier. Les théories cognitives classiques tentent de résoudre cette question en réduisant la signification à une description syntaxique, logique et propositionnelle. Celle-ci est donc relativement arbitraire et imposée de l'extérieur, si l'on ose dire, au système cognitif. Les théories connexionnistes, en mettant entre parenthèses le problème de l'émergence de la signification, retrouvent le confort méthodologique du behaviorisme d'antan, mais ne contribuent pas plus que le cognitivisme orthodoxe à la solution du problème sémantique (Shanon, 1992 ; Tiberghien, 1996). A moins, bien sûr, d'admettre que les états locaux codent eux-mêmes des symboles - mais, dans ce cas, la frontière entre les deux classes de théories s'estompe fortement et le connexionnisme devient lui-même représentationnel (Hinton, 1986 ; Rumelhart & Todd, 1993, pp. 13-21). On comprend que plusieurs chercheurs tentent, malgré les difficultés méthodologiques et théoriques considérables, de développer des systèmes hybrides, symbolico-connexionnistes. Dans ce contexte, les théories représentationnelles et connexionnistes décriraient, de façon non contradictoire, des niveaux d'organisation différents: processus de haut niveau gouvernés par des règles, pour les premières, et processus perceptifs, de type classification de patterns, pour les secondes (Anderson, 1990 ; Smolensky, Legendre, & Myata, 1993).

Toutes ces difficultés peuvent expliquer l'apparition, dans les années 80, d'un programme qui rejette toute orthodoxie, qu'elle soit représentationnaliste ou connexionniste. Ce mouvement radical postule que la cognition peut être adéquatement décrite sous la forme d'un réseau. Cependant, pour qu'un tel réseau puisse produire de la signification, il doit nécessairement posséder une histoire, agir sur son environnement et être sensible aux variations contextuelles de cet environnement. D'ailleurs cette distinction entre système cognitif et environnement doit, elle même, être réévaluée et le concept de représentation peut aussi être mis radicalement en cause. Ainsi, le monde n'est pas préformé et n'est pas progressivement représenté, le monde et l'action sont co-déterminés. C'est l'historique des actions en contexte qui fait émerger un monde de significations, un monde "énacté", pour reprendre l'expression de Varela (1988, 1996). Si le connexionnisme peut apparaître, rétrospectivement, comme une revanche posthume de la première cybernétique, on devine aussi tout ce que le programme radical de l'énaction doit à la seconde cybernétique, celle de Foerster, Ashby et Grey-Walter.

### **L'avenir de la psychologie cognitive ... et celui des sciences cognitives**

La contribution de la psychologie scientifique à la naissance des sciences cognitives a donc été déterminante. Le behaviorisme tolmanien et la gestalttheorie, en tentant de surmonter les contradictions du behaviorisme orthodoxe, ont préparé l'émergence des théories représentationnelles. L'épistémologie génétique de Piaget, en proposant une description axiomatisée de l'intelligence, a été une des sources importantes des théories formelles de la connaissance dans les sciences cognitives. Le néo-behaviorisme hullien et la mathématisation croissante de la psychologie (modèles de mesure psychophysique et modélisation stochastique de l'apprentissage) ont préparé la voie au programme connexionniste. Le connexionniste radical, lui-même, reprend d'ailleurs de nombreuses hypothèses formulées par la gestalttheorie, la psychologie topologique et la psychologie écologique. A partir des années 70, l'histoire de la psychologie cognitive devient inséparable de celle des sciences cognitives. Elle se caractérise par des transferts permanents entre les différentes composantes des sciences de la cognition: transferts méthodologiques (analyse chronométrique des temps de réaction, oculométrie, méthodes psychophysiques, paradigmes d'amorçage, technique de dissociation de processus, par exemple), transferts empiriques (loi de la spécificité de l'encodage, empan mnémonique, effet de prototypie, fonction précision-rapidité, par exemple), transferts théoriques (distinction épisodique/sémantique, déclaratif/procédural, implicite/explicite, par exemple). L'intégration des recherches devient si grande que d'aucuns pourraient être tentés de prédire la fin de la psychologie cognitive -- et peut-être même la fin de la psychologie ! Si l'on fait provisoirement abstraction des contraintes institutionnelles, l'avenir de la psychologie cognitive dépendra, en grande partie, de la solution qui sera apportée à plusieurs questions méta-théoriques. Trois de ces questions ont, me semble-t-il, une portée épistémologique critique, mais sans doute n'est-ce au fond qu'une seule et même question. Il s'agit, tout d'abord, des implications de l'hypothèse selon laquelle la cognition est organisée de façon modulaire. La deuxième question est lancinante et porte sur la relation entre la cognition et le cerveau. La troisième est celle du rapport de la philosophie aux sciences cognitives, en général, et à la psychologie cognitive, en particulier.

### **La modularité de l'esprit et l'idéologie modulariste**

L'hypothèse de la modularité de l'esprit, défendue par Fodor en 1983 établit que tout stimulus distal est transformé par des systèmes "subsidiaries" (perceptuels) en un ensemble de signaux neuronaux analogiques qui peuvent être traités par des systèmes "périphériques". Ces derniers réalisent des inférences calculatoires obligatoires, spécifiques et non



conscientes. Plusieurs critères permettent, en principe, d'identifier ces systèmes périphériques: 1) ils sont propres à un domaine (perception de la couleur, de la forme, des visages, etc.) ; 2) leurs opérations sont obligatoires et ne peuvent être contrôlées intentionnellement ; 3) ils sont rapides ; 4) ils sont informationnellement cloisonnés et insensibles aux influences latérales des autres systèmes périphériques et, a fortiori, aux influences de plus haut niveau; 5) leurs sorties sont superficielles ; 6) ils s'inscrivent physiquement dans des architectures neuronales fixes ; 7) ils présentent des défaillances spécifiques ; 8) ils présentent une ontogenèse caractéristique (Fodor, 1983/1986). Les sorties de ces systèmes périphériques sont combinées et intégrées dans les systèmes "centraux" analogiques et non spécifiques. Ils mettent en jeu, simultanément, plusieurs sorties des systèmes périphériques (isotropie computationnelle). Leurs opérations sont inscrites dans de nombreux circuits neuronaux distincts (isotropie neuronale). Ils permettent les inférences rationnelles, la fixation de croyance, la mémoire et, plus généralement, la pensée. En résumé, l'hypothèse de la modularité rompt radicalement avec la théorie holistique de l'esprit et repose sur une architecture séquentielle et ascendante ("bottom-up") pour décrire le flux des traitements cognitifs.

Cette hypothèse doit être distinguée de ce que j'appellerai l'idéologie "modulariste", sorte de pensée unique d'une espèce de cognitivisme post-moderne. Elle repose sur certains propos imprudents et/ou provocateurs tenus par Fodor et que l'on peut résumer sous la forme de "lois", les trois "lois de Fodor": 1) "Plus un processus est global, moins on sait de choses à son sujet" (1983/1986, p. 140) ; 2) "il n'existe pas de psychologie scientifique sérieuse des processus centraux" (1983/1986, p. 166) ; 3) "Si l'on ne sait rien de la neuropsychologie de la pensée, c'est qu'il n'y a rien à savoir sur la neuropsychologie de la pensée" (1983/1986, p. 153). L'adhésion à cette conception signifierait incontestablement la fin de la psychologie cognitive puisque celle-ci étudie les systèmes périphériques mais n'exclut pas la possibilité d'une exploration des systèmes centraux. Se limiter à l'étude des seuls systèmes périphériques faciliterait incontestablement la réduction, sans conditions, de la psychologie aux neurosciences (cognitives), voire à la "neuroscience cognitive". Mais cette théologie cognitive, normative et autoritaire, est scientifiquement indéfendable. Certes, on connaît en général moins de choses sur les processus globaux (qui sont, en général, plus complexes) que sur les processus locaux (qui sont, en général, plus simples) mais il n'en découle, ni logiquement ni empiriquement, que les seconds nous sont toujours connaissables tandis que les premiers nous resteraient à jamais opaques. Il y a d'ailleurs d'importantes différences entre le niveau de connaissance que nous possédons sur les divers systèmes périphériques dans des domaines différents (la motricité, la vision, l'olfaction, etc.). Affirmer qu'il n'y a aucune psychologie sérieuse des processus centraux témoigne surtout d'une méconnaissance de l'histoire de la psychologie scientifique. Les recherches sur la mémoire humaine sont un contre-exemple évident de cette affirmation : c'est un des domaines de la psychologie qui a sans doute le plus progressé au cours des deux dernières décennies. Les résultats obtenus et les théories de la mémoire ont d'ailleurs contribué très largement au développement même de la neuropsychologie (Schacter & Tulving, 1994 ; Tiberghien, 1997). D'ailleurs, au cours des dernières années, les nouvelles technologies des neurosciences (méthode lésionnelle, imagerie cérébrale, potentiels évoqués, etc.) ont confirmé de nombreuses lois empiriques et validé plusieurs théories élaborées, dans ce domaine, par la psychologie cognitive.

Il en va, bien sûr, autrement de l'hypothèse de la modularité. Elle doit être évidemment prise au sérieux et évaluée très précisément. La valeur heuristique de cette hypothèse est d'ailleurs incontestable. Elle a engendré une stratégie de recherche particulièrement efficace fondée sur la dissociation des processus cognitifs. Elle a également permis une convergence rapide entre les recherches en psychologie cognitive et celles conduites en neuropsychologie et en neurosciences. Et pourtant, dans sa formulation fodorienne initiale, elle est sans aucun

doute fausse ou, à tout le moins, beaucoup trop simplificatrice pour rendre compte de la complexité des observations expérimentales ou cliniques.

•  
Considérons, à titre d'illustration, la reconnaissance des visages dont la nature modulaire a été explicitement supposée par Fodor (1983/1986, p. 68). Si l'on accepte la conjecture fodorienne, on devrait observer que l'ensemble des critères de la modularité s'appliquent à la reconnaissance des visages. Or la question est très controversée. Tout d'abord la définition de ce domaine pose problème puisque la reconnaissance du visage implique nécessairement un traitement d'informations très élémentaires (forme, contour, contraste, couleur, etc.) et d'informations beaucoup plus complexes (angle de vue, genre, expression, identité, etc.). En d'autres termes, la reconnaissance met en oeuvre une "cascade" de traitements, de complexité variable, modulés par l'orientation et le niveau attentionnel ainsi que par le contexte général de l'activité et le type de décision requis. Bien sûr, on observe effectivement, dans le cortex occipito-temporal, des ensembles de neurones qui répondent très spécifiquement aux visages et à leurs propriétés. Mais on constate que ces neurones "faciaux" sont situés dans des régions qui répondent aussi, dans des proportions variables, à d'autres stimuli et à une grande variété de dimensions élémentaires. Pour ne citer qu'un exemple, des neurones du cortex temporal du singe sont sensibles à l'identité et ne sont pas activés par des variations de l'expression faciale -- alors que des neurones sensibles à l'expression peuvent répondre aussi à l'identité (Hasselmo, Rolls, & Baylis, 1989 ; Perrett et al., 1984). De plus, les régions activées varient selon le type de traitement facial: la discrimination du genre active la surface ventrale du cortex temporal antérieur, l'identification du visage active une région plus antérieure et, enfin, la dénomination active une région plus antérieure encore. En d'autres termes tous les attributs visuels d'un objet ne sont pas stockés, de façon unitaire, dans une région définie du cortex, et leur traitement spécifique peut faire intervenir des régions différentes (Ungerleider, 1995, pp. 770-771 ; Tulving, 1995). Attirer l'attention sur ce point ne veut pas dire, on l'aura compris, qu'il n'y a pas de localisation fonctionnelle au niveau cérébral, ce qui serait contraire aux données empiriques, mais cela remet en cause une conception souvent trop étroite du cloisonnement informationnel .

•  
La neuropsychologie de la reconnaissance des visages confirme largement les conclusions neuro-anatomiques. Il existe, par exemple, un déficit neuropsychologique relativement rare, mais spectaculaire, qui se traduit par une incapacité à reconnaître les visages familiers: la prosopagnosie ou agnosie visuelle des visages. Ce trouble est associé à des lésions le plus souvent bilatérales, d'étendue variable, du gyrus occipito-temporal (gyrus fusiforme, gyrus lingual et gyrus parahippocampique). On pourrait considérer que cette défaillance spécifique, impliquant une architecture neuronale relativement fixe satisfait, en apparence, à au moins deux des critères essentiels de la modularité fodorienne. Une telle conclusion serait, à tout le moins, prématurée. En effet, selon les cas décrits dans la littérature, les caractéristiques des processus cognitifs perturbés sont très variables. Certains patients ont essentiellement des troubles perceptifs, de type métamorphosies, qui empêchent toute discrimination entre les visages ; d'autres présentent surtout des troubles catégoriels (du genre, de la race, de l'âge) de telle sorte que tous les visages paraissent similaires ; d'autres présentent principalement des troubles de l'estimation de la familiarité induisant un échec de la reconnaissance; enfin, d'autres ressentent normalement la familiarité mais sont incapables d'identifier la personne correspondant à un visage (Ellis, 1986). Si l'on se limite à un traitement particulier, l'expression faciale, on constate que certains patients prosopagnosiques ont un problème à percevoir et à catégoriser les expressions faciales alors que d'autres n'en ont pas (Davidoff & Landis, 1990 ; Sansone, 1994 ; Schweich & Bruyer, 1993). Ainsi, bien que les régions cérébrales impliquées dans la prosopagnosie soient assez précisément délimitées, les

processus cognitifs perturbés et leur architecture sont encore loin d'être élucidés. A tel point que l'on parle prudemment de formes différentes de prosopagnosies dont l'origine peut être perceptive ou mnésique et, dans ce dernier cas, consécutives à un dysfonctionnement des processus de la représentation ou de l'accessibilité (McNeil & Warrington, 1991 ; De Renzi, 1986) . Comme le déclarait, très prudemment, Fodor (1983/1986, p. 130): "Tout mécanisme psychologiquement distinct (...) peut subir en principe (souligné par nous, GT) un handicap sélectif". On ne saurait mieux dire.

Ainsi une conception modulaire stricte des processus périphériques est difficilement compatible avec une analyse rigoureuse des données empiriques. Mais la conception fodorienne des systèmes centraux s'avère elle-même difficilement tenable. En effet certains processus centraux ne sont pas strictement isotropes et ont des propriétés qui les rapprochent des systèmes périphériques. Les recherches en neuropsychologie ont montré que certains domaines, comme la mémoire sémantique, sont associés à des architectures neuronales fixes. C'est ainsi que certaines lésions cérébrales entraînent un déficit, connu sous le nom d'anomie. Cette incapacité à se rappeler le nom des objets peut se révéler très spécifique et n'affecter que la dénomination des êtres vivants sans aucun déficit de la dénomination des objets artificiels (Warrington & Shallice, 1984). De nombreux déficits sélectifs de la mémoire sémantique ont ainsi été observés : noms abstraits vs noms concrets, noms propres vs noms communs, etc.(Chertkow, 1990 ; Lucchelli & De Renzi, 1992 ; Warrington, 1981). On a aussi invoqué, de façon sans doute encore plus spéculative, l'hypothèse fodorienne de "domaine-spécificité" pour expliquer le développement ontogénétique de la catégorisation et de la représentation des concepts de haut niveau. Une telle hypothèse supposerait des principes d'organisation innés permettant de sélectionner, à un âge très précoce, des ensembles d'attributs corrélés afin de catégoriser l'environnement de façon pertinente (Pacherie, 1996 ; Spelke, 1990).

En définitive, que reste-t-il de la théorie générale de Fodor ? Plus grand chose de sa pureté originelle et intransigeante. En effet, elle reposait sur un postulat essentiel opposant les systèmes périphériques aux systèmes centraux. Or les critères de définition des systèmes périphériques s'appliquent très inégalement, et souvent de façon partielle, aux différents niveaux de traitement et aux différents domaines d'information -- et ceci même pour les plus élémentaires. Pire, les systèmes dits centraux présentent également des propriétés qu'ils partagent parfois avec des systèmes périphériques. La conscience elle-même pourrait même être conçue, à la limite, comme modulaire (Tiberghien & Jeannerod, 1995, p. 183). Et, si tout est modulaire, comment est alors réalisée l'intégration de ces différentes sources d'information et les inférences logiques qui peuvent en découler ? Enfin, seuls les trois derniers critères de Fodor permettent de différencier clairement un module fodorien d'une simple capacité sur-apprise. On comprend, dans ces conditions, que Baron-Cohen préfère parler de mécanismes neurocognitifs, au lieu de modules, afin d'expliquer la nature pré-spécifiée du processus de lecture mentale dont le dysfonctionnement pourrait être à l'origine de l'autisme (Baron-Cohen, 1995/1998, p. 72). Faut-il donc maintenir cette dichotomie fodorienne dont la logique, si elle était poussée jusqu'à son terme, aboutirait à interdire tout programme de recherche en psychologie cognitive et, sans doute, en sciences cognitives ? Evidemment non. Tout converge, au contraire, vers l'hypothèse plus complexe d'un continuum de traitements, organisés selon une architecture hautement interactive et très probablement en cascade, possédant à des degrés divers et sous des contraintes plus ou moins fortes, tel ou tel sous-ensemble des propriétés modulaires. S'il y a des modules, et surtout des quasi-modules, à différents niveaux de l'organisation cognitive, alors l'étude des systèmes centraux peut être raisonnablement envisagée.

La psychologie cognitive ne peut donc survivre si l'on accepte purement et simplement la logique fodorienne. Sous sa forme originelle, elle interdit l'étude des processus cognitifs centraux, dits de haut niveau. Sous sa forme amendée, elle réduit les processus centraux à des modules périphériques. Dans les deux cas ces postulats aboutissent, ou risquent d'aboutir, à un réductionnisme neurologique intégral. En effet, si tout processus cognitif est de type modulariste et si un module admet nécessairement une architecture neuronale fixe, les sciences cognitives se confondront, à plus ou moins long terme, avec les neurosciences. Celles-ci phagocyteraient alors inéluctablement la psychologie cognitive. Seules demeurerait les neurosciences cognitives et la cognition sera réduite aux seuls phénomènes susceptibles d'être décrits en termes de modules. Le modularisme fodorien ne se contente donc pas de dessaisir le centre au profit de la périphérie, il finit aussi par réduire le centre ... à du périphérique ! Le behaviorisme avait été beaucoup moins loin puisqu'il n'avait jamais nié les processus centraux, il les avait simplement mis provisoirement entre parenthèses.

### **Le réductionnisme neurobiologique et la nature des sciences cognitives**

La place et le devenir de la psychologie cognitive sont donc largement dépendants de la réponse que l'on peut apporter à la conjecture fodorienne. Mais derrière ce débat se dissimule un autre problème, encore plus fondamental, et qui en explique probablement l'intensité. Il s'agit en effet de savoir si les descriptions psychologiques de la cognition peuvent être progressivement réduites à des descriptions neurobiologiques. En d'autres termes, peut-on "naturaliser" l'esprit ?

Cette question du réductionnisme neurobiologique n'est évidemment pas nouvelle mais les progrès récents de la neuropsychologie et, en particulier, des techniques d'imagerie cérébrale lui ont donné encore plus d'acuité. Ils ont aussi renforcé la croyance selon laquelle c'est une description de type neurobiologique, et non de type psychologique, qui permettrait d'expliquer la nature et les propriétés de la cognition. Autrement dit, ce seraient les neurosciences qui "illumineraient" la psychologie. Je suis d'un avis exactement opposé et je l'illustrerai de deux façons différentes.

Et tout d'abord, qu'apporte précisément l'observation des lésions cérébrales, provoquées ou observées, à la compréhension des phénomènes cognitifs. C'est évidemment une des pièces centrales de la discussion de l'hypothèse réductionniste. Plus précisément, est-il possible d'inférer le fonctionnement cognitif à partir des seules observations lésionnelles en neuropsychologie? Pour répondre positivement à une telle question, il faut accepter au moins deux méta-hypothèses, celle du "fractionnement" et celle de la "transparence" (Caramazza, 1986, 1992). L'hypothèse du "fractionnement" suppose qu'un processus cognitif "normal" (P) implique un l'état du cerveau "normal" (M) et qu'un processus cognitif "pathologique" (P\*) implique un état du cerveau affecté par une lésion identifiable (L) :  $P \rightarrow M$  et  $P^* \rightarrow M+L$ . La première de ces hypothèses a trois implications : 1) elle suppose que le cerveau est cognitivement non homogène ; 2) elle admet un effet spécifique des lésions sur les processus cognitifs ; 3) enfin, elle considère que la méthodologie expérimentale de dissociation de comportements, classique en neuropsychologie, revient à dissocier des processus cognitifs. La seconde de ces hypothèses a également plusieurs implications: 1) elle suppose que le cerveau est un système modulaire, au sens fodorien du terme ; 2) elle admet une conception strictement additive des processus cognitifs ; 3) elle considère qu'une lésion déterminée n'agit que sur un seul processus (ou sous-processus) cognitif ; 4) elle reconnaît que les conséquences d'une lésion sont irréversibles.

L'hypothèse de fractionnement et celle de transparence sont classiquement admises en neuropsychologie mais elles ne vont pas de soi (pour une critique : Kosslyn & Intriligator, 1992 ; Morton, 1992). En effet, si la validité générale de l'hypothèse de fractionnement est confirmée par le rôle fonctionnel de nombreuses régions cérébrales, elle soulève toutefois de nombreuses difficultés. Il est clair, tout d'abord, que la localisation de régions du cerveau, spécialisées dans tel ou tel traitement cognitif, ne peut être décrite indépendamment des nombreuses interconnexions qui les relient . La neurologie a d'ailleurs beaucoup plus progressé dans l'étude de ces localisations que dans celle des interconnexions dont l'importance fonctionnelle est sans doute décisive. De plus l'effet des lésions varie sensiblement, à la fois selon leur étendue et selon les techniques (neurotoxiques ou par aspiration, par exemple) qui ont été mises en oeuvre pour les produire chez l'animal. Des lésions définies peuvent avoir des effets non spécifiques et des lésions diffuses peuvent avoir des effets caractéristiques (Sansone, 1994). Enfin, on sait aussi que de nombreux problèmes méthodologiques ne permettent pas de passer facilement, et directement, d'une dissociation comportementale à une dissociation de processus (Jacoby, 1993 ; Neely, 1989). L'hypothèse de la transparence peut également être critiquée car on dispose maintenant de nombreuses données psychologiques et neurologiques qui démontrent la non-linéarité du fonctionnement cérébral et la non-additivité des processus cognitifs (Globus, 1992). La réalisation de n'importe quelle tâche cognitive met nécessairement en oeuvre de nombreux processus cognitifs en interaction constante. Enfin plusieurs études de cas en neuropsychologie illustrent clairement la possibilité de stratégies compensatoires chez certains patients cérébrolésés (Bruyer, 1991).

L'étude des localisations fonctionnelles cérébrales par l'observation des lésions spontanées, chez l'homme, ou provoquées, chez l'animal, est donc à elle seule insuffisante pour reconstruire conceptuellement la cognition humaine. Comme le fait remarquer Gabrieli (1998, p. 88), le comportement des patients atteints de lésions cérébrales ne nous renseigne pas sur les processus perturbés, mais plutôt sur ce que leur cerveau peut encore réaliser en dépit même de la lésion. En définitive, cela revient à émettre un sérieux doute sur l'hypothèse de transparence. Mais l'hypothèse du fractionnement est tout aussi problématique et l'exploration des effets des seules lésions cérébrales ne peut apporter de connaissances directes sur le fonctionnement du cerveau normal.

D'autres techniques d'exploration du cerveau pourraient prétendre passer de l'autre côté du miroir de la cognition. Nous pensons ici à l'enregistrement de l'activité cellulaire unitaire et à la neuro-imagerie (IRMf et TEP). Les premières permettent de déterminer les propriétés neuronales associées aux processus cognitifs normaux mais, malheureusement, elles ne permettent pas de localiser la source cérébrale exacte de ces propriétés. Les secondes permettent de "visualiser" les processus cognitifs mais elles n'offrent qu'une mesure indirecte de l'activité neuronale, puisque celle-ci est médiatisée par des variations du débit sanguin ou du métabolisme cérébral, ce qui en limite fortement la sensibilité spatiale et temporelle. La méthode de soustraction, classique dans ce domaine d'investigation, ne permet pas de savoir si l'absence de différence entre une image contrôle et une image expérimentale reflète des processus cognitifs de même nature ; elle permet uniquement d'affirmer que l'activité métabolique sous-jacente est identique. D'ailleurs des champs cérébraux pourraient participer à des opérations cognitives identiques mais présenter néanmoins des activations différentes. On ne peut donc, à partir d'une simple localisation d'un champ et de la mesure d'une intensité d'activation, en déduire la nature des processus cognitifs sous-jacents (Roland, 1995, p. 783). Une stratégie de complémentarité devrait donc s'imposer afin de réduire ce véritable principe d'incertitude cognitive où la fonction cognitive et les mécanismes neuronaux qui l'assurent ne peuvent être "mesurés" dans le même temps, avec la même précision et avec la même méthode.

On comprend que l'évolution récente des sciences cognitives ait pu apparaître à certains observateurs comme le triomphe du réductionnisme neurophysiologique. Je pense qu'ils se trompent. Ce n'est pas la psychologie cognitive qui est absorbée par les neurosciences, c'est à mon sens exactement le contraire: ce sont les neurosciences qui deviennent de plus en plus cognitives, qui ne peuvent plus faire l'économie du concept de cognition et nier l'importance de la modélisation cognitive. La neuropsychologie et les neurosciences cognitives passent une très grande partie de leur temps, peut-être même la totalité, à rechercher la localisation cérébrale des fonctions cognitives élaborées dans les laboratoires de psychologie cognitive. Qu'étudie-t-on, par exemple, en neuro-imagerie de la mémoire, sinon les modèles et théories élaborés en psychologie cognitive, certains depuis une vingtaine d'années: la mémoire épisodique active-t-elle des régions cérébrales distinctes de celles qui le sont par la mémoire sémantique ? La distinction entre le déclaratif et le procédural est-elle inscrite dans le cerveau ? Les représentations conscientes (explicites) mettent-elles en oeuvre les mêmes systèmes neuronaux que les représentations non conscientes (implicites) ? Les processus d'encodage et de récupération en mémoire activent-ils les mêmes réseaux de neurones ? etc. (Nyberg, Cabeza, & Tulving, 1996 ; Tulving, 1995). De ce point de vue, certaines expressions sont trompeuses. Parler d'imagerie cérébrale ne peut avoir, au fond, qu'un sens métaphorique car il n'y a pas à proprement parler, d'images du cerveau. Les images dont il s'agit sont des images "cognitives", des images de la cognition qui n'ont de signification que si l'on dispose d'un modèle cognitif. L'imagerie cérébrale est, en fait et en principe, une imagerie cognitive et l'importance de cette "image" là c'est qu'elle est à l'interface du monde physique et du monde mental. C'est non seulement une nouveauté, mais sans doute une révolution méthodologique. On ne produit plus seulement des "schémas" du cerveau mais des "cartes" du cerveau qui, tout comme les cartes géographiques, renvoient à une entité qui est à la fois inscrite dans la structure physique de la carte mais aussi en dehors d'elle, "référéncée" par elle.

Cette métaphore de la carte topographique peut d'ailleurs servir aussi à illustrer la distinction entre déduction et explication: l'agencement des formes et des couleurs inscrites sur une carte peut être déduite, en principe, de sa structure physico-chimique mais celle-ci ne peut suffire à les interpréter, c'est-à-dire, en définitive, à les expliquer. De même, on peut, en principe, déduire la macrostructure neuronale (MN) de la microstructure physico-chimique du cerveau (MPC). Puis, toujours en principe, déduire les états globaux du cerveau (C) de MPC et, enfin, les comportements (R) -- voire les états mentaux (M) -- à partir de C. Cette relation déductive est purement descriptive et strictement transitive : MPC -->MN-->C-->M-->R. Elle permet, en principe des déductions du type MPC -->R ou MPC-->M. L'erreur du réductionnisme est de confondre cette relation déductive avec une relation explicative: MPC==>MN==>C==>M==>R. La relation explicative, à la différence de la première, n'est pas transitive et n'implique pas que l'on puisse expliquer le comportement ou les états mentaux à partir de la microstructure physico chimique du cerveau : MPC ==>R ou MPC==>M. De même que la carte ne peut-être expliquée sans référence avec le "terrain" et de nombreuses hypothèses auxiliaires (codage symbolique, par exemple), la cognition ne peut être expliquée sans référence à un environnement (physique mais aussi social et culturel) et un historique (Putnam, 1973 ; Putnam, 1988/1990).

### **Philosophie cognitive et antipsychologisme néo-kantien**

L'avenir de la psychologie cognitive dans les sciences cognitives dépend, en partie, des relations proprement scientifiques qu'elle entretient avec d'autres disciplines et, en premier lieu, avec les neurosciences. Mais son avenir dépendra aussi de la façon dont elle va poser et résoudre un certain nombre de questions de nature épistémologique. Nous avons évoqué deux de ces questions: la première porte sur les propriétés computationnelles de l'esprit et

nous l'avons discutée à partir de l'hypothèse fodorienne de la modularité ; la seconde porte sur le problème de l'inscription cérébrale de l'esprit et nous l'avons discutée à partir du réductionnisme neurophysiologique. Le fait même que la psychologie scientifique soit, une nouvelle fois, confrontée à la philosophie, et qu'elle le soit précisément au sein des sciences cognitives, constitue un paradoxe des plus intéressants. En effet nous avons déjà évoqué la difficile émancipation de la psychologie scientifique à l'égard de la philosophie en général et de la psychologie philosophique en particulier. La psychologie scientifique, behavioriste et réflexologique, a été fondamentalement empiriste. Les rapports avec la philosophie portèrent, essentiellement, pendant près d'un siècle, sur la question de l'objectivité (de la vérification) et sur le débat récurrent opposant l'idéalisme au matérialisme. (Le Ny, 1970). La psychologie cognitive, à ses débuts, n'a pas franchement rompu avec cette position épistémologique. Comme l'écrivait Piaget, vers le milieu des années 60, la différence essentielle entre la psychologie scientifique et la psychologie philosophique ne tient pas à la différence entre les faits et les essences, ni à la notion d'intentionnalité, ni à celle de signification, ni même à l'utilisation de l'introspection mais à la possibilité ou non d'un contrôle expérimental (Piaget, 1965).

Toutefois, la fin du behaviorisme et l'essor de la psychologie cognitive vont modifier peu à peu les rapports de la psychologie scientifique à la philosophie. L'empirisme fonctionnaliste, philosophie spontanée des psychologues, va sensiblement reculer au profit d'un empirisme logique, voire même d'un nouveau rationalisme encouragé par l'important développement de la modélisation mathématique en psychologie et par le succès de la métaphore computationnelle. La psychologie cognitive mais plus encore, bien sûr, les sciences cognitives, allaient replacer au centre des recherches les concepts de représentation, de signification et d'intentionnalité. Ce fut "pain béni" pour la philosophie en général et pour la philosophie "de l'esprit" en particulier, véritable philosophie "cognitive" qui pouvait, en un second "round" décisif, prendre une revanche épistémique éclatante sur la psychologie scientifique et sur l'autonomie qu'elle avait revendiquée à son égard vers la fin du siècle dernier. Bien sûr il y a aussi des philosophes qui souhaitent une interaction équilibrée entre psychologie scientifique et philosophie dans le creuset des sciences cognitives (Engel, 1994 ; Pacherie, 1993). Mais on ne peut tout à fait oublier ici que la philosophie de l'esprit est née précisément au moment de la rupture entre la psychologie et la philosophie et que ses fondateurs rejetaient radicalement l'approche psychologique. Une philosophie de l'esprit "aprioriste" ou une "neurophilosophie" de l'esprit revendiquent bien, au fond, la fin de la psychologie scientifique. En effet, la première l'oblige à reconnaître qu'elle n'est qu'une simple variété de psychologie du sens commun, une psychologie "ordinaire", c'est-à-dire relevant "logiquement" -- épistémologiquement -- de la description philosophique. La seconde l'oblige à reconnaître qu'elle n'est rien car son objet n'existe plus, ou doit être étudié par d'autres, et qu'elle sera donc par conséquent "éliminée".

On aura reconnu, sous une présentation volontairement polémique et légèrement détournée, le débat sur la psychologie "populaire" ("folk psychology") et le dilemme lancinant entre le "représentationnalisme" et "l'éliminativisme". Les théories qui se réclament du représentationnalisme (Fodor, Marr, Neisser, Pylyshyn, Simon, etc.) posent une série de sept postulats: 1) la machine de Turing et l'ordinateur Von Neumann offrent un modèle approprié de la cognition; 2) le modèle de la cognition est de nature logique (logique propositionnelle) ; 3) les processus cognitifs sont de type descendant ("top-down") et il n'y a pas de computation sans représentation mentale ; 4) les niveaux de traitement (computationnel, algorithmique, physique) sont indépendants ; 5) il y a des états mentaux (cognitifs) autonomes par rapport aux états neuronaux ; 6) ces états mentaux peuvent être décrits sous la forme de symboles, de représentations linguistiques ou de règles et ils sont

structurés comme un langage ("mentalais") ; 7) enfin, la psychologie "populaire" est une théorie proto-scientifique susceptible d'être évaluée objectivement et de servir de base aux sciences cognitives. Les théories éliminativistes (P.M. & P.S. Churchland, Sejnowski, Smolenski, etc.) s'opposent, sur tous les points, aux théories précédentes : 1) la métaphore informatique de la cognition n'est pas appropriée ; 2) le modèle de la cognition ne s'inscrit pas dans un formalisme logique mais dans un réseau connexionniste ; 3) les processus cognitifs sont de type ascendant et les computations n'impliquent pas de représentations ; 4) il y a interdépendance entre les niveaux de traitement; 5) il convient d'adhérer au réductionnisme neurophysiologique ; 6) il n'y a ni représentations mentales ni langage de la pensée ; 7) la psychologie populaire ne peut servir de base scientifique aux sciences cognitives.

Si la psychologie scientifique est sommée de choisir entre ces deux positions extrêmes, son avenir devient alors tout à fait problématique. Le piège épistémologique est particulièrement efficace et redoutable car, quel que soit le choix, il risque d'annoncer la fin de la psychologie scientifique. En effet l'option éliminativiste aboutit inéluctablement à la réduction totale et irréversible de l'objet de la psychologie à celui des neurosciences. A mon sens, les conséquences d'un tel choix seraient strictement identiques pour toute philosophie qui adhérerait sans réserve au programme éliminativiste : elle se transformerait inéluctablement en une «philosophie des neurones», provisoirement appelée neurophilosophie, mais simple appendice épistémologique des neurosciences, aux antipodes du programme de la philosophie de l'esprit. Même si les risques semblent moins évidents, le choix du représentationnalisme est tout aussi dangereux pour l'avenir de la psychologie cognitive expérimentale. Si la psychologie n'est qu'une forme «savante» de la psychologie populaire, son objet d'étude peut être aisément annexé par une forme «populaire» de philosophie savante, en l'occurrence la philosophie de l'esprit. La psychologie cognitive n'aurait-elle d'autre choix que de se dissoudre dans une philosophie cognitive ou d'être finalement réduite à une neuroscience cognitive ? Evidemment non, et tout ceci n'est qu'un rêve «éveillé» cartésien, mais sous sa forme la plus cauchemardesque.

On peut refuser résolument le choix proposé entre ces deux épistémologies prototypiques (Engel, 1991, p. 145). Pourquoi ne pas admettre un niveau de description représentationnel, celui de la psychologie populaire, et un niveau de description qui en spécifie les propriétés physiques et neuronales ? Pourquoi ne pas accepter une interaction entre flux de traitement ascendant et descendant ? Ne peut-on, à la fois, refuser le réductionnisme neuronal et ne pas accepter pour autant l'indépendance totale des niveaux de traitement, les plus complexes pouvant être décrits en termes de logique mentale (pas forcément la logique représentationnelle) et les plus élémentaires en termes connexionnistes ? Représentationnalistes et éliminativistes commettent, en quelque sorte, la même erreur, mais de façon complémentaire. Les premiers expliquent le comportement par les représentations mentales mais, ne découvrant pas la base physique qui permet de les déduire, ils en arrivent à affirmer que les symboles sont de nature physique et ont donc, en eux-mêmes, un pouvoir causal. Les éliminativistes ne parvenant pas à déduire les représentations mentales d'une simple description neuronale en arrivent alors à la conclusion que les états neuronaux expliquent causalement les états mentaux -- ces derniers étant dépourvus de toute efficacité causale. En d'autres termes, les premiers ne parviennent pas à déduire ce qu'ils peuvent, en principe, expliquer tandis que les seconds ne parviennent pas à expliquer ce qu'ils prétendent, en principe, déduire. Je suggère de décrire cette situation épistémique désagréable sous le nom de «paradoxe de Putnam».

Un choix épistémique aussi extrême peut être refusé et l'on peut très bien admettre la réalité



des états intentionnels de la psychologie populaire sans pour autant supposer qu'ils sont physiquement inscrits dans le cerveau. Il suffirait peut-être que ne soit maintenue constante que la seule structure de similarité entre le niveau des représentations symboliques et le niveau des représentations neuronales (Rumelhart & Todd, 1993, pp. 22-28). Au fond, c'est peut-être dans une épistémologie hybride que pourrait se trouver la solution pratique de cette lancinante question. En effet les modèles hybrides, ou symbolico-connexionnistes, pourraient permettre de dépasser la contradiction en construisant des architectures cognitives qui intègrent différents composants souvent considérés comme incompatibles : représentation symbolique et non symbolique, représentation distribuée et localisée, représentation abstraite et contextualisée, traitement parallèle et séquentiel, processus de haut niveau et de bas niveau. Cette stratégie de recherche est d'ailleurs encouragée par les nombreux résultats expérimentaux qui montrent que certains processus cognitifs sont en fait rapides, automatiques et "incoercibles" tandis que d'autres sont lents, délibérés et contrôlés. Les modèles hybrides de la cognition permettent de relier le niveau logico-symbolique de la psychologie populaire au niveau sub-symbolique des réseaux de neurones. Cette épistémologie "relativiste", mais "complémentariste", devrait être au centre des sciences cognitives en permettant à plusieurs langages de description et d'explication de coexister et de se renforcer mutuellement.

## Conclusion

Nous avons essayé de montrer que la psychologie cognitive n'est pas née après la première cybernétique, comme certaines chroniques pourraient le laisser entendre, mais bien avant. Son histoire se confond avec celle des sciences cognitives parce qu'elle en est à la fois l'origine et le centre. Survivra-t-elle aux sciences cognitives ? On peut le pronostiquer, mais à la condition qu'elle réussisse à éviter plusieurs écueils. Elle disparaîtrait, à coup sûr, si elle cédait devant à ce que j'ai appelé l'idéologie modulariste. En effet, dans ce cas, elle risquerait de ne se limiter qu'à l'étude de processus de très bas niveau, informationnellement cloisonnés, qui peuvent être, en principe, expliqués par une simple description neuro-comportementale. Pour se développer, la psychologie doit rejeter résolument cette idéologie et affirmer que les processus centraux non modulaires, comme l'attention ou la conscience par exemple, peuvent aussi être l'objet d'une enquête proprement scientifique. Et, comme l'a écrit judicieusement Gardner, "le temps travaille pour l'approche top-down" (1985/1993, p. 136).

La psychologie cognitive devrait également refuser le réductionnisme neurophysiologique et contester le concept même de "neuroscience cognitive". Les neurosciences cognitives ne sont évidemment pas en cause ici car il s'agit d'un des domaines les plus actifs et les plus prometteurs des sciences cognitives. Toutefois, promouvoir la neuroscience cognitive, comme programme épistémique, est fondamentalement réducteur et risque non seulement de mettre en danger la psychologie cognitive mais aussi les sciences cognitives elles-mêmes. Nous avons vu que la forme la plus actuelle, et la plus pernicieuse, de ce réductionnisme est actuellement un réductionnisme méthodologique. Subjugués par l'imagerie cérébrale, certains psychologues de la cognition en viennent à considérer qu'il n'y aura plus, dans un avenir proche, de recherche cognitive sans images du cerveau. Ils ont peut-être raison mais pas dans le sens où ils l'entendent. Nous avons essayé de montrer qu'il n'y avait pas d'images du cerveau à proprement parler ou, plus précisément, que ces images ne prennent leur sens qu'en raison même des théories et hypothèses cognitives élaborées par la psychologie cognitive. Contrairement aux apparences, spectaculaires et médiatiques, c'est la psychologie cognitive qui donne "esprit" aux images métaboliques du cerveau et non l'inverse. Autrement dit, c'est la psychologie cognitive qui vient aux neurosciences et non la psychologie cognitive qui se réduit progressivement aux neurosciences. Réduire un objet

d'étude, la cognition, aux indicateurs auxquels le chercheur a accès, à un moment donné, serait aussi préjudiciable aux sciences cognitives qu'à la psychologie cognitive. Pour ne citer qu'un exemple, il ne viendrait à l'idée de personne de contester l'importance considérable de la chronométrie mentale dans l'exploration des phénomènes cognitifs. Mais affirmer qu'il n'y a pas de psychologie cognitive sans enregistrement de la latence du comportement serait un cas de réductionnisme méthodologique aussi patent que celui que nous évoquions précédemment à propos de l'imagerie cérébrale. Il aurait d'ailleurs la même conséquence, celle d'interdire l'étude des phénomènes de haut niveau. C'est, en réalité, la combinaison de différents indicateurs qui nous fait progresser dans notre connaissance intime des processus cognitifs. Les neurophysiologistes le savent d'ailleurs très bien et ils sont souvent, paradoxalement, moins réducteurs que certains psychologues de la cognition séduits par l'esthétique et la simplicité apparente du diaporama cérébral. Mais ces indicateurs ne s'observent pas, ils s'inventent, et le moteur ultime de cette "invention" est constitué précisément par les théories et les modèles cognitifs. C'est pour cette raison que la psychologie cognitive ne devrait pas disparaître et que l'on peut même s'attendre à une intensification sensible des relations scientifiques qu'elle nouera avec les neurosciences cognitives dans le cadre, de plus en plus nécessaire, des sciences cognitives. Toutefois, il n'en demeure pas moins que des contraintes historiques, institutionnelles et sociales particulières -- en France, par exemple -- pourraient aboutir, localement, à un scénario beaucoup moins consensuel où les neurosciences finiraient par "dévorer" la psychologie cognitive.

## NOTES

1 Je me rappelle ici d'un Conseil d'Université, en 1981, où je défendais le programme des sciences cognitives. Mon plaidoyer me valut cette réplique définitive de la part d'un de mes éminents confrères: "La science cognitive ? Mais il s'agit là, mon cher collègue, d'une pure tautologie !". A l'époque une telle appréciation a certainement dû me contrarier. Mais, à la réflexion, cette façon de caractériser les sciences cognitives possédait une part insoupçonnée de vérité puisqu'il s'agit finalement de connaître notre propre connaissance ou, plus précisément, "de construire de la connaissance à propos de la cognition de tous" (Le Ny, 1989, pp. 28-29).

2 D'ailleurs, même si l'on choisit de situer l'origine des sciences cognitives lors de la première cybernétique, il n'est pas indifférent d'insister sur le rôle fondateur de tel ou tel événement scientifique et sur la place occupée par telle ou telle discipline. Se référer, par exemple, au Symposium Hixon (1948) permet de souligner le rôle décisif de l'interaction entre l'informatique et la neurobiologie dans la naissance des sciences cognitives. Invoquer les conférences de la Fondation Macy (1946-1953) conduira à accorder sans doute plus d'importance aux relations entre l'informatique et la psychologie. Les psychologues ont été très nombreux à participer aux débats de cette période (Bavelas, Boring, Harrower, Klüver, Köhler, Lewin, Marquis, Teuber, etc.), plus nombreux même que les cybernéticiens. L'influence de la psychologie cognitive dans les Conférences Macy deviendra même prépondérante à partir de 1950, alors que les cybernéticiens ne voyaient en elle "qu'un continent à conquérir", selon l'expression de Dupuy (1994, p. 48).

3 Affirmer, par exemple, que le cognitivisme est "incarné par le linguiste américain Noam Chomsky" (Fottorino, 1998, p. 13) est une généralisation historiquement discutable car elle fait l'impasse sur d'autres "incarnations" du cognitivisme toutes aussi importantes (Miller, Neisser, Piaget, Simon, Tolman, pour n'en citer que quelques unes).

4 On trouvera un historique des relations entre biologie et psychologie, du point de vue de la neurophysiologie, dans l'étude des origines de la physiologie mentale proposée par

Jeannerod (1996, pp. 136-147).

5 Il faut remarquer toutefois que le cognitivisme n'élimine pas, purement et simplement, le paradigme behavioriste selon lequel le comportement (R) est une fonction des seuls stimulus (S) observables dans l'environnement externe ou le milieu corporel. En réalité, le paradigme cognitiviste assimile, au sens le plus fort du terme, le schéma S-R et l'intègre à son propre paradigme dont il ne constitue qu'un sous-schéma emboîté. En effet le schéma théorique du cognitivisme implique que les comportements et les états mentaux (M) sont une fonction interactive de la variation des stimuli de l'environnement et d'autres états mentaux :  $R$  (ou  $M$ ) =  $f(S \times M')$ . D'ailleurs, les paradigmes de la psychologie introspective ( $M = f(M')$ ), de la psychophysique fechnérienne ( $R = f(M)$ ) et du néo-behaviorisme médiationnel ( $R = f(S \times S')$ ) peuvent aussi être considérés comme emboîtés dans le schéma programmatique du cognitivisme.

6 Quelques témoins de cette évolution peuvent être aujourd'hui clairement identifiés : publication, en 1951, de l'ouvrage d'Attneave présentant les applications à la psychologie de la théorie de l'information, élaborée par Shannon (Attneave, 1959) ; publication du premier manuel de psychologie cognitive "Cognitive psychology" (Neisser, 1967) ; première livraison de la revue "Cognitive Psychology" (1970) ; première livraison de la revue "Cognition" (1971).

7 Tolman décrit ainsi l'apprentissage animal en termes d'intention ("purpose") : tendances comportementales persistantes, sélectives et optimales orientées vers l'atteinte d'un but. Tolman admet que certains aspects du comportement animal (hésitations comportementales ou essais et erreurs vicariants, apprentissage latent, par exemple) sont accompagnés de "quelque chose de mental" que l'on peut décrire sous ce nom d'intention. Dans ce cas, le comportement exprime, selon lui, une véritable représentation cognitive ("cognitive postulation") (Tolman, 1925a, 1925b, 1927).

8 Quelques repères utiles dans l'histoire des sciences cognitives : origine du programme des sciences cognitives (Conférence de Cambridge, 1956) ; première apparition de l'expression "sciences cognitives" en 1970 ; création du "Journal of Cognitive Science" en 1976 ; Rapport Sloan aux USA pour le développement des sciences cognitives en 1978 ; enfin, création de la "Cognitive Science Society" en 1979.

9 Mais, évidemment, sur une machine séquentielle, comme le TEDVAC de Von Neumann, et non sur un ordinateur parallèle, comme l'ENIAC de McCulloch. Ce choix illustre clairement la rupture, à cette époque, entre la cybernétique et le cognitivisme, en psychologie comme en informatique.

10 Le programme de l'Intelligence Artificielle (IA), en rupture avec la cybernétique, domine la Conférence de Dartmouth durant l'été 1956 (y participent : Chomsky, McCarthy, Miller, Minsky, Simon). Il est symptomatique de remarquer que l'IA et les Sciences Cognitives naissent la même année, à quelques mois d'intervalle.

11 La modélisation stochastique a popularisé, en psychologie, l'utilisation du calcul numérique et probabiliste pour l'étude des phénomènes cognitifs (apprentissage et mémoire associatives). Le rôle catalyseur du C.H.I.P. ("Center for Human Information and Processing", La Jolla CA), dans les années 70, ne saurait être sous-estimé. En 1971, Rumelhart y modélise déjà la mémoire à long terme, en collaboration avec Lindsay et Norman. L'Institut des Sciences Cognitives, qui lui succède bientôt sera le creuset de la nouvelle approche connexionniste de la cognition. Ce fragment d'histoire est un bon indicateur de l'évolution théorique qui a permis le passage de la computation symbolique à la computation connexionniste. En 1979, McClelland publie un important article théorique précisant les principes méthodologiques et théoriques permettant l'analyse de processus cognitifs qui ne sont pas organisés séquentiellement mais en cascade. Le groupe P.D.P. ("Parallel Distributed Processing") voit aussi le jour en 1981, avec Rumelhart et McClelland,

dans le même environnement californien de l'Institut des Sciences Cognitives. On y retrouve Rumelhart, McClelland et, aussi, Norman, Sejnowski, Smolensky, etc. Le groupe P.D.P. est au connexionnisme ce que la Conférence de Cambridge a été aux sciences cognitives "représentationnelles" et il renoue, d'un certain point de vue, avec l'esprit des Conférences de Macy et de Hixon.

12 Le behaviorisme de Hull peut aussi être considéré comme un des précurseurs du programme connexionniste en sciences cognitives. Il entrevoit, dès les années 40, que l'apprentissage peut s'expliquer par l'interaction, en réseau, de plusieurs influx afférents modulés par des oscillations, excitatrices ou inhibitrices, d'origine centrale. Comme le fait remarquer Hintzman: "... les psychologues hulliens n'ont pas inventé nos modernes systèmes PDP (Parallel Distributed Processing) mais ils auraient pu le faire s'ils avaient disposés de stations de travail Sun ou d'un Cray" (1993, p. 375).

13 L'influence de la gestalttheorie est patente, par exemple, dans le modèle connexionniste FACADE (Grossberg, 1997). Ce modèle non linéaire permet de formuler des prédictions sur les mécanismes neuronaux susceptibles de rendre compte de la distinction compétitive entre figure et fond sur la base de contrastes de luminance ou de signaux de couleur. La théorie topologique du champ ("life space") de K. Lewin est aussi une belle intuition de l'application de la mécanique statistique à la modélisation cognitive (analogie entre le modèle de Hopfield et la théorie du champ magnétique, machine de Boltzmann, par exemple) (Hertz, Anders, & Palmer, 1991). La psychologie gibsonienne, avec son rejet total des concepts de représentation mentale et de computation, a contribué à ébranler les théories représentationnelles de l'esprit (Gibson, 1979). Si le concept "d'affordance" (actions potentielles existantes dans un objet ou une scène) peut trouver des résonances dans le connexionnisme classique, sa conception du "tout dans l'environnement" le démarque également fortement du courant connexionniste, surtout radical. La psychologie écologique est une tentative de rapprochement de ces deux programmes des sciences cognitives (Neisser, 1994).

14 Mettre en évidence des localisations fonctionnelles cérébrales est évidemment de la plus haute importance théorique. A condition toutefois de ne pas sous-estimer l'importance, au moins égale, des relations entre les régions fonctionnelles identifiées. Pour ne prendre qu'une illustration, il est classique d'opposer la voie dorso-pariétale (spécialisée dans la localisation spatiale) de la voie ventro-temporale (spécialisée dans l'identification des objets). Mais il ne faudrait pas oublier qu'il y a de nombreuses interactions entre ces deux flux d'informations -- par exemple, entre l'aire médio-temporale supérieure (dorsale) et l'aire inféro-temporale (ventrale), mais il y en a beaucoup d'autres (Ungerleider, 1995).

De même si l'architecture neuronale de ce système peut être dite "globalement" séquentielle puisque le flux des traitements "progresses", dans l'ensemble, des régions postérieures aux régions antérieures du cerveau. Mais elle ne l'est certainement pas de façon stricte. Presque toutes les relations entre les aires des deux systèmes, ventral et dorsal, sont bidirectionnelles et les effets descendants sont la règle plutôt que l'exception. Le fonctionnement n'est donc décrit, de façon adéquate, ni par un modèle séquentiel strict ni par un modèle parallèle. L'organisation des aires corticales du système visuel n'est ni sérielle ni parallèle, mais sans doute en cascade (McClelland, 1979). L'ordre d'activation ne suit pas forcément l'ordre de la description neuro-anatomique et il y a une grande simultanéité de traitements. Il n'y a pas seulement un traitement ascendant (feedforward) mais aussi des rétroactions (top-down) qui déterminent les propriétés des champs récepteurs et, au delà, de l'ensemble de la perception (Bullier & Nowak, 1995).

15 Et encore, à condition d'accepter l'hypothèse selon laquelle les formes observées de prosopagnosies résultent bien d'un dysfonctionnement propre au domaine des visages, ce qui est le premier des critères fodorien de la modularité. Cela peut aussi être mis en doute. On a observé, par exemple, un cas de prosopagnosie chez un ornithologue qui présentait non

seulement un déficit de la reconnaissance des visages mais aussi de celle des différentes espèces d'oiseaux (Bornstein, Sroka, & Munitz, 1969). Un cas de zooagnosie, associé à une prosopagnosie, a également été décrit (Assal, Favre, & Anderes, 1984). Rien ne permet donc d'écarter l'hypothèse selon laquelle les lésions qui sont à l'origine de la prosopagnosie affectent, en réalité, tout type d'expertise visuelle de niveau équivalent à celui requis pour reconnaître les visages. Les données expérimentales ayant montré que l'inversion spatiale perturbe beaucoup plus la reconnaissance du visage que celle des autres objets sont aussi en faveur de cet effet d'expertise. Celle-ci s'observe aussi pour des stimuli non faciaux mais seulement s'ils peuvent être discriminés de façon experte (Bruyer, 1992 ; Carey, 1992 ; Diamond & Carey, 1986 ; Gauthier & Tar, 1995). En d'autres termes, il n'est pas certain que l'expertise nécessaire au traitement des visages soit "domaine-spécifique", au sens fodorien du terme (Damasio, Tranel, & Damasio, 1990 ; Farah, 1994 ; Farah, Wilson, Drain, & Tanaka, 1998).

16 L'architecture "massivement" connexionniste et non-additive du cerveau explique d'ailleurs que l'effet d'une lésion puisse ne pas être nécessairement perturbateur. Par exemple, une lésion de l'hippocampe peut atténuer un déficit de la mémoire de reconnaissance visuelle induit par une lésion du cortex périrhinal (Meunier, Hadfield, Bachevalier, & Murray, 1996).

17 Le concept de "psychologie populaire " est un concept central de la philosophie de l'esprit selon lequel des attitudes propositionnelles (consistant à attribuer des croyances, désirs, souhaits, etc.) ont un contenu intentionnel ou sémantique. Ce dernier possède un pouvoir causal, et donc prédictif, sur le comportement d'un individu.

18 On peut distinguer quatre classes principales de modèles hybrides: les systèmes basés sur l'interaction entre un module connexionniste et un module symbolique ; les systèmes principalement symboliques mais avec des composants locaux neuronaux ; les systèmes entièrement connexionnistes permettant une émergence de symboles à partir des interactions neurales ; les réseaux connexionnistes localisés avec attribution arbitraire de symboles à certaines unités (Tiberghien, 1997, pp. 150-155).

19 Ils savent ainsi, par exemple, que les résultats des études lésionnelles sont souvent sur-interprétés et qu'ils ne nous renseignent pas sur les mécanismes cognitifs sous-jacents. Ils savent encore que des phénomènes cérébraux sont, momentanément ou durablement, accessibles à certaines méthodes d'exploration ... mais pas à d'autres. Cela a été le cas assez longtemps, par exemple, pour de nombreuses techniques d'imagerie cérébrale qui ne montraient pas, ou peu, d'implication de l'hippocampe dans l'encodage et la récupération mnésiques alors que les méthodes pharmacologiques démontraient le contraire (Izquierdo & Medina, 1998).

## REFERENCES BIBLIOGRAPHIQUES

- Anderson, J. A. (1984). Cognitive psychology. *Artificial Intelligence*, 23, 1-11.
- Anderson, J. A. (1990). Hybrid computation in cognitive science: Neural networks and symbols. *Applied Cognitive Psychology*, 4, 337-347.
- Andler, D. (Ed.). (1992). *Introduction aux sciences cognitives*. Paris: Gallimard.
- Assal, G., Favre, C., & Anderes, J. P. (1984). Non-reconnaissance d'animaux familiers chez un paysan: zoo-agnosie ou prosopagnosie pour les animaux. *Revue Neurologique*, 140, 580-584.
- Attneave, F. (1959). *Applications of information theory to psychology: A summary of basic concepts, methods and results*. New York: Holt, Rinehart and winston.
- Baron-Cohen, S. (1998). *La cécité mentale. Un essai sur l'autisme et la théorie de l'esprit* (J. Nadel. & F. Lefebvre., Trans.). Grenoble: Presses Universitaires de Grenoble. (Ouvrage original publié en 1995.).

- Bornstein, B., Sroka, H., & Munitz, H. (1969). Prosopagnosia with animal face agnosia. *Cognition*, 5, 164-169.
- Bruyer, R. (1991). Covert face recognition in prosopagnosia: A review. *Brain and Cognition*, 15, 223-235.
- Bruyer, R., Crispeels, G. (1992). Expertise in person recognition. *Bulletin of the Psychonomic Society*, 30(6), 501-504.
- Bullier, J., & Nowak, G. (1995). Parallel versus serial processing: New vistas on the distributed organization of the visual system. *Current Opinion in Neurobiology*, 5, 497-503.
- Caramazza, A. (1986). On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance : The case for single-patients studies. *Brain and Cognition*, 5, 41-66.
- Caramazza, A. (1992). Is cognitive neuropsychology possible ? *Journal of Cognitive Neuroscience*, 4(1), 80-95.
- Carey, S. (1992). Becoming a face expert. *Philosophical Transactions of the Royal Society, London*, B335, 95-103.
- Chertkow, H., Bub, D. (1990). Semantic memory loss in dementia of the Alzheimer type. *Brain*, 113, 397-417.
- Damasio, A. R., Tranel, D., & Damasio, H. (1990). Face agnosia and the neural substrates of memory. *Annual Review of Neuroscience*, 13, 89-109.
- Davidoff, J., & Landis, T. (1990). Recognition of unfamiliar faces in prosopagnosia. *Neuropsychologia*, 28, 1143-1161.
- De Renzi, E. (1986). Current issues in prosopagnosia. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. Young (Eds.), *Aspects of face processing* (pp. 243-252). Dordrecht: Martinus Nijhoff.
- Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General*, 115(2), 107-117.
- Dupuy, J.-P. (1994). *Aux origines des sciences cognitives*. Paris: La Découverte.
- Dupuy, J. P. (1985). L'essor de la première cybernétique (1943-1953). *Cahiers du C.R.E.A.*, 7(Novembre), 7-140.
- Ellis, H. D. (1986). Disorders of face recognition. In K. Poeck, Freund, H.J., Gänshirt, H. (Ed.), *Neurology* (pp. 179-187). Berlin: Springer Verlag.
- Engel, P. (1991). Psychologie populaire et explication cognitive. In J. N. Missa (Ed.), *Philosophie de l'esprit et sciences du cerveau* (pp. 135-146). Paris: Librairie Philosophique Vrin.
- Engel, P. (1994). *Introduction à la philosophie de l'esprit*. Paris: La Découverte.
- Estes, W. K. (1960). Learning theory and the new "mental chemistry". *Psychological Review*, 67, 207-223.
- Farah, M. J. (1994). Neuropsychological inference with an interactive brain: A critique of the "locality" assumption. *Behavioral and Brain Sciences*, 17(1), 43-104.
- Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception ? *Psychological Review*, 105(3), 482-498.
- Fodor, J. A. (1986). *La modularité de l'esprit: essai sur la psychologie des facultés* (A. Gerschenfeld, Transl.). Paris: Editions de Minuit. (Ouvrage original publié en 1983)
- Fodor, J. A., Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3-71.
- Fottorino, E. (1998, Mardi 3 Février). Voyage au centre du cerveau: 1. Un monde imaginé. *Le Monde*, pp. 12-13.
- Gabrieli, J. D. E. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology*, 49, 87-115.
- Gardner, H. (1993). *Histoire de la révolution cognitive: La nouvelle science de l'esprit* (J-L. Peytavin, J-L., Trans.). Paris: Payot. (Ouvrage original publié en 1985)
- Gauthier, I., & Tar, M. J. (1995, November). Becoming a "greeble" expert: exploring the face recognition memory. Paper presented at the Psychonomic Society, Los Angelès.

- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.
- Globus, G. G. (1992). Toward a noncomputational cognitive neuroscience. *Journal of Cognitive Neuroscience*, 4, 299-310.
- Goussard, M., Grégoire, P., Matalon, B., & Piaget, J. (1959). *La logique des apprentissages, Etudes d'épistémologie génétique* (Vol. 10, ). Paris: Presses Universitaires de France.
- Grossberg, S. (1997). Cortical dynamics of three-dimensional figure-ground perception of two-dimensional pictures. *Psychological Review*, 104(3), 618-658.
- Hasselmo, M. E., Rolls, D. I., & Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioral Brain Research*, 32, 203-218.
- Hertz, J., Anders, K., & Palmer, R. G. (1991). *Introduction to the theory of neural computation*. (Vol. I). Redwood City, CA: Addison-Wesley Publishing Company.
- Hinton, G. E. (1986). Learning distributed representations of concepts. Paper presented at the Cognitive Science Society, Amherst.
- Hinton, G. E., Anderson, J.A. (1989). *Parallel models of associative memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hintzman, D. L. (1993). Twenty-five years of learning and memory: Was the cognitive revolution a mistake ? In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 359-391). Cambridge, MA: The MIT Press.
- Hull, C. L. (1943). *Principles of Behavior: An introduction to behavior theory*. New York: Appleton-Century-Crofts.
- Hull, C. L. (1952). Fractional antedating goal reactions, A behavior system (pp. 124-155). New Haven: Yale University Press.
- Jacoby, L. L., Toth, J.P., Yonelinas, A.P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, 122, 139-154.
- Jeannerod, M. (1996). *De la physiologie mentale : Histoire des relations entre biologie et psychologie*. Paris: Odile Jacob.
- Kosslyn, S. M., & Intriligator, J. M. (1992). Is cognitive neuropsychology plausible? The perils of sitting on one-legged stool. *Journal of Cognitive Neuroscience*, 4, 96-106.
- Le Ny, J.-F. (1970). *Psychologie et matérialisme dialectique*. Bruxelles: Editions Le Pavillon.
- Le Ny, J.-F. (1989). *Science cognitive et compréhension du langage*. Paris: Presses Universitaires de France.
- Lucchelli, F., & De Renzi, E. (1992). Proper name anomia. *Cognition*, 28, 221-230.
- McClelland, J. L. (1979). On the time relations of mental processes: an examination of systems of processes in cascade. *Psychological Review*, 86, 287-330.
- McClelland, J. L., Rumelhart, D.E. and the PDP research group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Vol. 2: Psychological and biological models. (Vol. 2). Cambridge, MA: The MIT Press.
- McNeil, J. E., & Warrington, E. K. (1991). Prosopagnosia: a reclassification. *Quarterly Journal of Experimental Psychology*, 43A, 267-287.
- Meunier, M., Hadfield, W., Bachevalier, J., & Murray, E. A. (1996). Effects of rhinal cortex lesions combined with hippocampectomy on visual recognition memory in rhesus monkeys. *Journal of Neurophysiology*, 75, 1190-1205.
- Morton, J., & Marshall, J.C. (1992, 24-28 August). Seven questions in cognitive neuroscience. Paper presented at the Trieste Encounter on "The interaction between cognitive and biological science" (SISSA), Trieste.
- Neely, J. H. (1989). Experimental dissociations and the episodic/semantic memory distinctions. In I. Roediger, H.L. & F. I. M. Craik (Eds.), *Varieties of memory and consciousness : Essays in honour of Endel Tulving* (pp. 229-270). Hillsdale, NJ: Lawrence

Erlbaum Associates.

Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Neisser, U. (1994). Multiple systems: A new approach to cognitive theory. *European Journal of Cognitive psychology*, 6(3), 225-241.

Nyberg, L., Cabeza, R., & Tulving, E. (1996). PET studies of encoding and retrieval: The HERA model. *Psychonomic Bulletin & Review*, 3(2), 135-148.

Osgood, C. E. (1960). *Method and theory in experimental psychology*. New York: Oxford University Press.

Pacherie, E. (1993). *Naturaliser l'intentionnalité : Essai de philosophie de la psychologie*. Paris: Presses Universitaires de France.

Pacherie, E. (1996). Domaines conceptuels et modularité. In V. Rialle & D. Fiset (Eds.), *Penser l'esprit: de la cognition à une philosophie cognitive* (pp. 175-185). Paris: Presses Universitaires de Grenoble.

Pélissier, A., & Tête, A. (1995). *Sciences cognitives : Textes fondateurs (1943-1950)*. Paris: Presses Universitaires de France.

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., & Jeeves, M. A. (1984). Neurons responsive to faces in the temporal cortex : Studies of functional organization, sensitivity to identity and relation to perception. *Human Neurobiology*, 3, 197-208.

Piaget, J. (1965). *Sagesse et illusions de la philosophie*. Paris: Presses Universitaires de France.

Putnam, H. (1973). Reductionism and the nature of psychology. *Cognition*, 2(1), 131-146.

Putnam, H. (1990). *Représentation et réalité*. (C. Engel-Tiercelin, Trans.). Paris: Gallimard. (ouvrage original publié en 1988).

Pylyshyn, Z. (1982). Litterature from cognitive psychology. *Artificial Intelligence*, 19, 251-255.

Roland, P. E., Kawashima, R., Gulyás R., & O'Sullivan, B. (1995). Positron Emission Tomography in cognitive neuroscience: Methodological constraints, strategies, and examples from learning and memory. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 781-788). Cambridge MA.: The MIT Press.

Rumelhart, D. E., & Todd, P. M. (1993). Learning and connectionist representations. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV : Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 3-30). Cambridge, MA: The MIT Press.

Sansone, S., Tiberghien, G. (1994). Traitement de l'expression faciale et reconnaissance des visages: Indépendance ou interaction ? *Psychologie Française*, 39(4), 327-344.

Schacter, D. L., & Tulving, E. (1994). *Memory systems 1994*. Cambridge, MA: Bradford Book & The MIT Press.

Schweich, M., & Bruyer, R. (1993). Heterogeneity in the cognitive manifestations of prosopagnosia: The study of a group of single cases. *Cognitive Neuropsychology*, 10, 529-547.

Shanon, B. (1992). Are connectionist models cognitive ? *Philosophical Psychology*, 5, 235-255.

Simon, H. A., & Kaplan, C. A. (1989). Foundations of cognitive science. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 1-47). Cambridge, MA: The MIT Press.

Smolensky, P., Legendre, G., & Myata, Y. (1993). Integrating connectionist and symbolic computation for the theory of language. *Current Science*, 64(6), 381-391.

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, 14, 29-55.

Spence, K. W. (1956). *Behavior theory and conditioning*. New Haven: Yale University Press.

Tiberghien, G. (1996). Le connexionnisme: Stade suprême du behaviorisme ? In V. Rialle & D. Fiset (Eds.), *Penser l'esprit: de la cognition à une philosophie cognitive* (pp. 27-41). Grenoble: Presses Universitaires de Grenoble.



- Tiberghien, G. (1997). *La mémoire oubliée*. Liège: Mardaga.
- Tiberghien, G., & Jeannerod, M. (1995). Pour la science cognitive: La métaphore cognitive est-elle scientifiquement fondée ? *Revue Internationale de Psychopathologie*, 18, 173-203.
- Tolman, E. C. (1925a). Behaviorism and purpose. *Journal of Philosophy*, 22, 36-41.
- Tolman, E. C. (1925b). Purpose and cognition: The determiners of animal learning. *Psychological Review*, 32, 285-297.
- Tolman, E. C. (1927). A behaviorist's definition of consciousness. *Psychological Review*, 34, 433-439.
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55, 189-208.
- Tulving, E. (1995). Organization of memory: Quo vadis ? In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 839-847). Cambridge MA.: The MIT Press.
- Ungerleider, L. (1995). Functional brain imaging studies of cortical mechanisms for memory. *Science*, 270, 769-775.
- Varela, F. (1988). *Les sciences cognitives: Tendances et perspectives actuelles*. Paris: Le Seuil.
- Varela, F. J. (1996). *Invitation aux sciences cognitives*. Paris: Seuil.
- Vergnaud, G. (1991, Juillet-Août). Pourquoi la psychologie cognitive ? *La Pensée*, 282, 9-19.
- Warrington, E. K. (1981). Concrete word dyslexia. *British Journal of Psychology*, 72, 175-196.
- Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *B*, 107, 829-854.
- White, M. J. (1983). Prominent publications in cognitive psychology. *Memory & Cognition*, 11 (4), 423-427.