

Article original

Entre neurosciences et neurophilosophie : la psychologie cognitive et les sciences cognitives

Between neuroscience and neurophilosophy: cognitive psychology and cognitive science

G. Tiberghien

Institut des sciences cognitives, 67, boulevard Pinel, 69675 Bron cedex, France

Reçu le 7 mars 2005 ; accepté le 5 mai 2007

Résumé

Les sciences cognitives sont à la fois un champ scientifique, un enjeu épistémologique et un lieu de confrontation institutionnelle entre plusieurs disciplines académiques. Dans ce contexte, la psychologie cognitive occupe une position historique et théorique centrale. Il est donc essentiel de situer précisément sa contribution à l'origine et au développement des sciences cognitives. Cependant, pour certains, la psychologie cognitive pourrait se fondre dans la neuroscience cognitive dont les techniques d'imagerie cérébrale permettraient, à terme, de dépasser l'obsédant problème du dualisme cerveau–esprit. Pour d'autres, un tel pari épistémologique ne peut être tenu qu'en réduisant la cognition à ses états et processus les plus élémentaires. Le problème de la représentation mentale (du sens et de la conscience) est alors invoqué pour exclure une conception étroitement réductionniste de l'esprit au cerveau. Les sciences cognitives ne seront-elles donc qu'une simple étape de l'intégration de la psychologie cognitive aux neurosciences cognitives ? Ou, au contraire, les autres disciplines constitutives des sciences cognitives seront-elles conduites à placer la psychologie cognitive au centre de leur programme de recherche et à en reconnaître la spécificité ? Le débat est en cours mais l'issue scientifique et institutionnelle incertaine.

© 2007 Société française de psychologie. Publié par Elsevier Masson SAS. Tous droits réservés.

Adresse e-mail : tiberghien@isc.cnrs.fr (G. Tiberghien).

Abstract

The cognitive sciences are all at once a scientific field with far-reaching epistemological implications and the locus of institutional confrontation between several academic disciplines. In this context, cognitive psychology occupies a key historical and theoretical position. It is therefore essential to clearly situate its contribution to the origin and the development of the cognitive sciences. For some, cognitive psychology may end up being dissolved by cognitive neuroscience, with its brain imaging techniques, which in the long run, should solve the age-old problem of brain-mind dualism. For others, this epistemic wager can only be won by reducing cognition to its most elementary states and processes. Here, the problem of mental representation (of meaning and consciousness) is brought to bear to rule out a view that strictly reduces the mind to the brain. Will the cognitive sciences, then, be nothing more than a mere stage in the integration of cognitive psychology into the cognitive neurosciences? Or on the contrary, will the other disciplines in the cognitive sciences be led to recognize the specificity of cognitive psychology and put it at the centre of their research program? The debate is underway, but the scientific and institutional outcome is uncertain.

© 2007 Société française de psychologie. Publié par Elsevier Masson SAS. Tous droits réservés.

Mots clés : Cerveau ; Cognition ; Esprit ; Neuro-imagerie ; Neurosciences cognitives ; Phrénologie ; Philosophie cognitive ; Psychologie cognitive ; Sciences cognitives

Keywords: Brain; Cognition; Cognitive neuroscience; Cognitive philosophy, Cognitive psychology; Cognitive science; Mind; Neuroimager; Phrenology

— *Mais vous n'avez plus de théorie, plus rien...*

— *Bianca, on ne peut disséquer l'esprit comme le cerveau !*

— *Mais toutes les recherches de neurophysiologie sur le fonctionnement du cortex, les cinq aires visuelles du cortex, vous n'y croyez plus ?*

— *C'est la sensation, Bianca. Pas la perception. Personne ne sait comment fonctionne mon cerveau lorsque je vous reconnais et vous trouve ravissante !*

— *Et le goût des zamponi ?*

— *C'est pareil.*

— *Mais enfin... il est possible de mettre en évidence par résonance magnétique nucléaire ou caméra à positons une augmentation d'activité de certaines aires corticales !*

— *Foutaises que tout cela, ma chère Bianca, on peut trouver ce que l'on cherche en modifiant le traitement statistique des données brutes. Et puis ce n'est pas nouveau ! Déjà Broussais en 1831 avait mesuré son crâne avec un craniomètre très perfectionné, avant et quatre ans après avoir été élu à l'Académie des sciences morales et politiques. Pendant cette période, il s'était livré à un travail intellectuel important. Eh bien, il avait constaté que sa « protubérance métaphysique » (selon la phrénologie de Gall) avait augmenté de trois millimètres !*

Michel Juvet, Le voleur de songes, (Juvet, 2004).

L'émergence des sciences cognitives, au milieu du siècle dernier, et leur structuration progressive au cours des décennies suivantes ont profondément modifié les contours de la psychologie. Elle a, de toute façon, complètement remanié l'ensemble des relations de la psychologie scientifique à la biologie, d'une part, et à la philosophie, d'autre part. Nous nous proposons,

dans cet article, de dégager les caractéristiques essentielles de cette évolution et de discuter deux questions qui nous semblent critiques pour l'avenir de la psychologie cognitive. Il s'agit, tout d'abord, de la relation causale entre les phénomènes neuronaux et les entités cognitives suggérées par la psychologie cognitive ; il s'agit, ensuite, de la contribution des techniques de neuro-imagerie à la solution de la question précédente.

Nous sommes, bien sûr, parfaitement conscients que cet historique est reconstruit à partir d'un certain point de vue, qui est celui de la psychologie cognitive. Parler d'un autre point de vue engendrerait sans doute un historique quelque peu différent. Nous voudrions toutefois montrer ici que seule l'approche pluricompetente des sciences cognitives peut aboutir à une authentique science de la cognition qui ne soit pas accaparée, et donc soumise, réduite ou appauvrie, par telle ou telle discipline, y compris la nôtre, et cela quels qu'en soient les mérites scientifiques.

1. Un ménage à trois plutôt tourmenté : biologie, philosophie et psychologie

1.1. De la psychophysiologie à la physiologie mentale

Les relations de la biologie à la psychologie scientifique sont anciennes et elles en ont même été, en grande partie, constitutives, pour des raisons à la fois de principe et de méthode. La méthode expérimentale, qui avait fait ses preuves dans le domaine biologique, allait en effet s'avérer aussi féconde pour l'étude psychophysique du psychisme et celui-ci sera mis très tôt en relation avec la totalité de l'organisme. La psychologie physiologique (ou psychophysiologie) visait, dès l'origine, à mettre en évidence d'éventuelles corrélations anatomopsychiques et psychosomatiques.

Mais cette entreprise a d'abord été conduite dans un cadre épistémologique relativement prudent et ouvert, et dans lequel, d'ailleurs, le corps avait autant d'importance que le système nerveux central. Ce cadre épistémologique était, de plus, davantage fondé sur le postulat d'un parallélisme psychophysiologique que sur celui d'un réductionnisme physiologique radical. Cette prudence transparaît très nettement, par exemple, chez James (1890/1891), l'un des fondateurs de la psychologie scientifique moderne, quand il déclare : « Le "spiritualiste" et "l'associationniste" doivent être tous les deux des "cérébralistes", dans la mesure où ils doivent *au moins* admettre que certaines propriétés de leurs lois préférées ne peuvent être expliquées que si les lois du cerveau en *codéterminent* les effets » (p. 4). Woodworth (1922/1930), auteur d'un manuel de psychologie très influent en son temps, est encore sur la même position quand il affirme : « ... toutes les activités psychologiques d'un être vivant résultent aussi de l'activité de ses organes et de ses cellules (...) Notre connaissance psychologique doit donc prendre en compte *au moins* un peu de physiologie (...) La psychologie physiologique est un effort pour *maintenir le contact* entre l'étude de l'individu et celle de ses organes et de ses cellules » (pp. 501,554).

Avec l'essor du béhaviorisme, les corrélations recherchées ont associé un nombre de plus en plus élevé de réflexes, puis de comportements ou de chaînes de comportements, à des états physiologiques du système nerveux, ou à des états organiques qui se trouvent sous son contrôle. La psychologie cognitive a ensuite permis d'établir des corrélations entre diverses entités cognitives, certaines très complexes comme la conscience par exemple, et des états physiologiques de l'organisme et du système nerveux. Mais, en tout état de cause, c'était toujours la psychologie qui était, dans ce contexte, considérée comme centrale.

Une autre ligne d'évolution a cependant abouti à inverser fondamentalement une telle logique et à considérer, cette fois, que c'est l'organisme et, surtout, le cerveau qui occupe la position centrale pour l'explication du comportement et, au-delà, de toute la cognition. C'est, bien sûr, la ligne qui a été suivie par la neurophysiologie, la neurologie et, dans une large mesure, par la neuropsychologie. Le passage aux neurosciences, comportementales puis cognitives, a encore accentué cette prééminence du cerveau sur l'esprit *mind*. C'est ce qu'affirment, par exemple et plutôt catégoriquement, Kosslyn et Koenig (1992) quand ils plaident pour une nouvelle neuroscience cognitive : « ... l'esprit est ce que fait le cerveau : une description de l'esprit est une description de la fonction du cerveau » (p. 4). L'essor des sciences cognitives et des nouvelles techniques d'exploration cérébrale n'a cependant pas créé les relations entre psychologie et (neuro)physiologie, elles étaient en fait souvent anciennes et approfondies. Elle a plutôt déplacé le centre de gravité des recherches, du psychique au neurologique et, aussi, du parallélisme psychophysiologique au réductionnisme neurobiologique. La psychologie physiologique a ainsi été transformée, par les sciences cognitives, en une sorte de physiologie mentale dont l'objectif, très clairement annoncé, est de naturaliser l'esprit. Mais, si le naturalisme ne se réduit pas, en principe, à un réductionnisme mécaniste, il peut le devenir dans la pratique scientifique, en partie socialement déterminée, et conduire progressivement à une intégration de la psychologie aux neurosciences.

On peut trouver une étude historique et théorique approfondie des rapports entre la psychologie et la biologie dans de nombreuses publications (Bonnet, 2000 ; Delay, 1945/1963 ; Dupéron, 2000 ; Fraisse et Piaget, 1966 ; Jeannerod, 1996, 1997 ; Piaget, 1963/1967 ; Richelle et Seron, 1994 ; Tiberghien, 1985a, 1985b). La lecture qu'en fait un neurophysiologiste est évidemment très instructive pour les psychologues de la cognition. Par exemple, pour Jeannerod (1996, p. 10), la psychologie n'est, au fond, qu'une discipline « d'attente » ou « de transition » qui sera peu à peu complètement intégrée aux sciences cognitives et devrait donc disparaître, à terme, en tant que discipline autonome. Un tel scénario n'est peut-être pas impossible, mais on devrait, logiquement, émettre le même pronostic pour les neurosciences et d'autres disciplines. Cependant, Jeannerod ne va pas jusque-là. Pourquoi ? C'est étrange, mais sans doute explicable si l'on suppose qu'un tel pronostic implique que les sciences cognitives auxquelles il est fait référence sont en réalité les seules neurosciences cognitives, c'est-à-dire des sciences cognitives intégrées, scientifiquement et institutionnellement, aux neurosciences.

Cette conception est d'ailleurs explicitée, sans détour cette fois, par Roy (2001) lorsqu'il affirme qu'il y a une dépendance, ontologique et logique, de la psychologie à la neurologie et que les sciences cognitives sont totalement intégrées aux neurosciences cognitives, celles-ci résultant de la « fusion » entre la psychologie cognitive et la neurobiologie. J'ai eu l'occasion de critiquer cette position dans plusieurs publications antérieures (Tiberghien, 1999 ; Tiberghien et Jeannerod, 1995) : si l'on veut éviter à la fois les pièges du dualisme et ceux du réductionnisme mécaniste, il est alors nécessaire de défendre une intégration disciplinaire dans un programme de recherche en sciences cognitives, et non dans un programme de recherche en neurosciences cognitives. Cela revient à soutenir que le processus d'émergence de la cognition à partir de ses causes neuronales, contraintes par l'évolution ontogénétique et phylogénétique, nécessite un niveau d'explication causale spécifique, qui est précisément celui qui doit être étudié par les sciences cognitives. Un programme naturaliste en sciences cognitives n'adhère donc pas nécessairement à un réductionnisme mécaniste (Decety, 1999).

1.2. Philosophie de la psychologie ou psychologie philosophique ?

Les relations entre la psychologie et la philosophie ont été d'une tout autre nature (pour une étude historique et théorique de ces rapports, Canguilhem, 1958 ; Piaget, 1965 ; Richard et Tiberghien, 1999). Si les recherches en physiologie, ont été, dès l'origine, une référence pour la psychologie scientifique, en revanche cette dernière s'est constituée, apparemment, en rupture complète avec la philosophie. Cela était sans doute inévitable car philosophie et psychologie chassaient, en partie, sur les mêmes terres et ces disciplines affirmaient, toutes deux, leur légitime prétention à expliquer le psychisme. Cette rupture a été à la fois théorique et méthodologique. Sur le plan théorique, l'antipsychologisme du rationalisme kantien a largement déterminé l'adhésion de la psychologie scientifique à un empirisme pratique et à un positivisme de principe. Sur le plan méthodologique, le divorce était inévitable entre l'introspection déductive, dérivée de l'anamnèse platonicienne, et l'inférence inductive de la méthode expérimentale.

Mais cette rupture épistémologique n'a jamais complètement interrompu les relations entre les deux disciplines et, peut-être même, dissimulait-elle un lien intime. Cela s'explique d'abord par des raisons sociohistoriques, car en Europe et surtout en France, la psychologie est restée très longtemps sous le contrôle institutionnel de la philosophie¹. Mais cela s'explique aussi du fait que tous les philosophes n'adhéraient pas au pessimisme kantien à l'égard de la psychologie. Enfin, les psychologues eux-mêmes étaient très largement influencés par les débats philosophiques, en particulier ceux portant sur le statut de la connaissance, et ceux-ci exerçaient une influence non négligeable sur certaines théories et pratiques d'investigation : l'essor de la théorie piagétienne, dite de l'épistémologie génétique, en est une belle illustration (Piaget, 1965, 1970) et sa contribution à la psychologie cognitive et aux sciences cognitives a été importante (pour une revue historique : Vauclair et Perret, 2003) ; de même, l'influence de l'épistémologie bachelardienne sur la mise en œuvre des méthodes quantitatives en psychologie est indéniable (Bachelard, 1965 ; Reuchlin, 1962, 1965). La psychologie a ainsi accédé très progressivement à une autonomie scientifique relative et à une indépendance institutionnelle complète à l'égard de la philosophie. Cette situation, qui se concrétise surtout après la Première Guerre mondiale (1939–1945), a peu à peu, et relativement, assagi les relations entre la philosophie et la psychologie.

Le développement des sciences cognitives, au cours des dernières décennies, a cependant très largement modifié la donne (pour une étude historique et théorique, voir : Dupuy, 1994 ; Engel, 1994 ; Gardner, 1985/1993 ; Pacherie, 1993 ; Richard et Tiberghien, 1999 ; Tiberghien, 1999). La philosophie analytique (ou philosophie de l'esprit, ou philosophie cognitive) rejette le psychologisme (voir Braunstein, ce numéro). De nouveau se pose avec acuité le problème des relations entre psychologie et philosophie, mais à propos, cette fois, de ce nouvel objet substantifié sous le nom de cognition. Dans le cadre des sciences cognitives, la philosophie de l'esprit ne se présente pas comme une simple épistémologie de la cognition, ce qui ne poserait pas de problème à la psychologie cognitive, mais elle prétend, au contraire, proposer une théorie complète de la cognition. En cela, elle se pose, de nouveau, en rival résolu de la psychologie scientifique, domaine qui, aujourd'hui, se confond largement, sinon totalement, avec celui de la psychologie cognitive. Le contrat, tacite, laborieusement établi entre philosophie et psychologie est donc rompu. Comme, en outre, la philosophie cognitive adhère résolument au

¹ Ce contrôle n'a d'ailleurs pas entièrement disparu. En effet, de nos jours encore, les seuls enseignements de psychologie dans le secondaire sont encore assurés, en France, par des professeurs de philosophie.

programme de la naturalisation de l'esprit, on assiste à une « nouvelle alliance », de fait, entre la philosophie et les neurosciences dont la psychologie cognitive pourrait bien être exclue, ou n'être acceptée que dans l'hypothèse de sa réduction complète ou de sa soumission, comme « champ intermédiaire », aux neurosciences cognitives (Netchine-Grynberg et Netchine, 1989).

La neurophilosophie illustre cette nouvelle alliance entre neurosciences et philosophie. Sous ce terme se manifeste une nouvelle science (sic) résultant de l'intégration des neurosciences et de la philosophie. Elle a été proclamée et défendue par Churchland (1986) qui considère que les théories fondées sur la psychologie populaire sont fausses, car elles reposent finalement sur des intuitions du sens commun, c'est-à-dire sur des contenus intentionnels. Mais, ce qui est valable pour la psychologie populaire l'est aussi pour la psychologie cognitive dont les théories postulent aussi, au moins en partie, des états représentationnels, donc intentionnels. En d'autres termes, l'éliminativisme cognitif de certains neurophilosophes est un réductionnisme qui élimine, à coup sûr, la psychologie cognitive comme programme scientifique autonome. Peut-on naturaliser la psychologie de cette façon sans, du même coup, la faire disparaître ou l'assujettir ? L'enjeu est de taille.

Un livre récent sur la philosophie cognitive (Pacherie et Proust, 2004) illustre, de façon symptomatique, l'asymétrie des relations que la philosophie entretient avec la psychologie et les neurosciences : l'un des chapitres est rédigé par un neurophysiologiste invité à débattre des rapports entre philosophie et neurosciences mais, curieusement, aucun psychologue de la cognition n'a été invité à débattre des rapports entre philosophie et psychologie ; pire encore, c'est un philosophe qui a rédigé ce chapitre dans lequel il utilise des données expérimentales obtenues en psychologie sociale cognitive afin de décider si la maïeutique platonicienne est une méthode d'inférence universelle ou relative (heureusement, pour la psychologie scientifique, elle est relative !). On ne voit pas très bien, dans ce contexte, pour quelle raison ce serait la philosophie cognitive qui délimiterait le champ empirique de la psychologie, en déciderait le degré de pertinence des données empiriques et, finalement, en proposerait les explications. La philosophie cognitive est-elle fondée à proposer « une théorie de la cognition qui rende compte de l'ensemble des modes de fixation des croyances, de la variété des savoir-faire, des théories implicites et des modes conscients ou non conscients d'accès à l'information » (Pacherie et Proust, 2004, p. 2) ? Je ne le crois pas, car la philosophie ne dispose pas des méthodes expérimentales qui lui permettraient d'élaborer de telles théories cognitives. C'est précisément la psychologie cognitive qui est en mesure de construire ces théories. Les sciences cognitives ne peuvent donc pas se réduire à un face-à-face privilégié entre la philosophie de l'esprit et les neurosciences cognitives.

2. Cerveau et cognition : si proches... et si lointains

Ce qui se trouve au centre des relations entre neurosciences cognitives, philosophie cognitive et psychologie cognitive c'est donc le problème du rapport entre le cerveau et la cognition. Il est évidemment impossible de contester, aujourd'hui, la causalité physicochimique propre aux phénomènes cérébraux. Les neurosciences ont, au cours de ces dernières décennies, largement progressé dans la description de ces événements cérébraux et dans l'explication de leurs relations causales. Ces événements cérébraux sont mis en relation avec des réponses comportementales, donc objectives, mais aussi avec des états cognitifs, conscients ou non conscients, postulés ou inférés. Il existe aussi, de façon symétrique, un rapport de causalité entre ces comportements, entre ces états cognitifs et entre comportements et états cognitifs.

Au cours du xx^e siècle, la psychologie comportementale et la psychologie cognitive ont largement contribué à explorer la nature de cette causalité propre à la génération des réponses comportementales et des états mentaux à partir de stimuli et de situations observables. Deux questions se posent alors :

- comment les états cognitifs sont-ils reliés aux états cérébraux ?
- Les états cognitifs peuvent-ils avoir une efficacité causale propre et quelle est alors la nature de cette efficacité causale ou, si l'on préfère, quelle est la nature de la relation entre ces deux niveaux de causalité, cérébral et cognitif ?

2.1. Comment la cognition est-elle causée par le cerveau ?

Il est impossible pour un scientifique, sauf à admettre une causalité divine ou magique, de refuser la proposition selon laquelle les états cognitifs sont causalement déterminés par les états du cerveau, la nature et l'origine de cette causalité ne pouvant résulter que de la phylogenèse, de l'ontogenèse et de l'histoire individuelle. Mais cette reconnaissance de principe n'est pas suffisante car elle n'implique pas obligatoirement qu'il soit possible pratiquement, par l'analyse et par l'expérimentation, de relier causalement l'état cognitif à l'état cérébral qui l'a produit (Putnam, 1973, 1988–1990). Il se pourrait, par exemple, que le nombre de microétapes causales entre l'état cérébral le plus élémentaire et l'état cognitif produit soit si élevé qu'il soit pratiquement impossible d'expliquer l'un par l'autre compte tenu, par exemple, de la non-transitivité explicative des pas élémentaires de déduction : soit une très longue chaîne causale $A \Rightarrow B \Rightarrow C \Rightarrow (\dots) \Rightarrow Y \Rightarrow Z$, le fait de pouvoir déduire localement B de A, C de B, ... Z de Y n'implique pas obligatoirement que l'on puisse déduire globalement Z de A et, par conséquent, expliquer Z par A.

Exprimé autrement, cela signifierait que le cerveau et l'esprit seraient un seul et même objet, mais dont on ne pourrait appréhender, dans le même temps et au même degré de précision, les deux constituants : une description précise du niveau cérébral, par exemple de type moléculaire, ne pouvant être mise en relation explicative qu'avec une description très simplifiée du niveau cognitif correspondant, par exemple en termes d'activation ou d'inhibition. De façon symétrique, une description très élaborée du niveau cognitif, par exemple en termes sémantiques ou conceptuels, ne pourrait être mise en relation explicative qu'avec un niveau de description relativement grossier au niveau cérébral correspondant, par exemple en termes de localisation neuroanatomique.

L'analogie avec le principe d'incertitude s'impose ici : les éléments du couple cerveau–esprit seraient aussi indissociables que ceux de la dyade espace–temps et l'on ne pourrait analyser, au même moment et au même niveau de profondeur et d'explication, l'un et l'autre de ces objets : gagner en précision sur un plan serait inéluctablement perdre sur l'autre (Bohr, 1932/1959/1961). De mon point de vue, ce sont bien évidemment les sciences cognitives qui apportent un indispensable principe de complémentarité : l'approche combinée de différentes disciplines, sous différents angles d'attaque et avec différentes méthodes et techniques d'étude, peut permettre d'approcher une description et une explication pertinente, quoique toujours approximative, de l'esprit–cerveau ou du cerveau–esprit. Les sciences cognitives n'étudient donc pas le cerveau, les neurosciences s'en chargent, ni l'esprit, la psychologie cognitive s'en charge. Les sciences cognitives étudient le couple cerveau–esprit ou, plus précisément, la

structure causale qui détermine l'esprit à partir du fonctionnement du cerveau et la façon dont ces deux entités sont liées en un seul objet par le sens (sur cette question : Jacob, 2000 ; Le Ny, 2005, pp. 15–16, 27–28).

2.2. Causalité cérébrale et causalité cognitive

La seconde question est de savoir si la nature causale des états cognitifs doit être distinguée de celle des états cérébraux et si elle possède une efficience propre. De nombreux psychologues et philosophes répondent positivement à cette question et postulent qu'il existe des représentations mentales dont le fonctionnement est autonome, au moins relativement, par rapport à leurs bases neurophysiologiques. Cette conception classique est connue sous le nom de *Théorie représentationnelle de l'esprit* (TRE). Elle postule :

- qu'un agent cognitif possède des états internes (mentaux) qui représentent le monde de façon plus ou moins véridique. C'est le problème de l'intentionnalité. Le monde représenté est donc un « monde » propre à l'agent cognitif mais dont la validité est, en principe, vérifiable. La pensée est alors décrite sous la forme de séquences d'états mentaux intentionnels ;
- que les représentations mentales du monde sont supposées complètes et durables, sinon permanentes ;
- que les représentations mentales du monde sont de nature symbolique ;
- que certaines de ces représentations mentales sont spécifiques et liées à des systèmes ancrés sur la perception (expérience) et d'autres sont amodales (abstraites). Ce postulat implique donc que la flexibilité cognitive résulte des capacités d'abstraction de l'agent cognitif et se trouve, par conséquent, sous le contrôle de processus de haut niveau *top-down* ;
- que la description de ces représentations mentales peut s'effectuer sans qu'il soit nécessaire de les mettre en relation avec les systèmes perceptivomoteurs de l'agent cognitif, c'est-à-dire avec le corps.

La nature symbolique des représentations mentales est incontestablement la pierre de touche de la TRE. Selon la TRE, la pensée est une manipulation de séquences d'états mentaux symboliques et intentionnels. Mais l'application systématique de la métaphore informatique a engendré une version plus moderne de la TRE, la *Théorie computationnelle de l'esprit* (TCE). Cette dernière adopte l'ensemble des postulats de base de la TRE, mais y ajoute un postulat additionnel :

- les symboles mentaux sont des séquences syntaxiquement organisées et sémantiquement évaluables, comme celles d'un programme informatique. Leur traitement résulte d'opérations de calcul (computations) selon les règles de la logique formelle.

Une alternative à cette TCE classique pourrait être la *Théorie computationnelle connexionniste de l'esprit* (TCCE). On a souvent présenté le connexionnisme comme une rupture radicale avec la TCE classique. Mais de nombreux modèles connexionnistes n'abandonnent pas le postulat de la nature symbolique des représentations. Des assemblées de neurones, naturels ou formels, peuvent très bien coder localement des représentations symboliques correspondant, par exemple, au sens des mots. Le connexionnisme classique, contrairement aux apparences, peut donc lui aussi être basé sur les concepts de computation et de représentation.

Beaucoup moins nombreux sont les chercheurs en sciences cognitives qui répondent négativement à la question de l'autonomie des états mentaux. Un connexionniste pourrait ainsi rejeter le postulat computosymbolique et devenir ce que les philosophes appellent un éliminativiste. Celui-ci décrirait alors probablement la cognition sous la forme d'un réseau de neurones conçu comme un système dynamique, auto-adaptatif et non linéaire, qui évolue vers des états attracteurs ou « répulseurs » selon une trajectoire probabiliste définie dans un espace multidimensionnel. C'est alors une nouvelle théorie, dite *Théorie dynamique de l'esprit* (TDE). On a même proposé le terme « d'enaction » (Varela, 1996) pour décrire cette cognition, inséparable du corps, qui ne « calcule » ni ne « représente », mais qui se comporte comme un système de résonance entre un esprit et un monde indissociables (un champ, au sens topologique du terme). La description de tels systèmes dynamiques est tout à fait compatible avec les propriétés connues du système nerveux. La question de leur validation empirique est autrement épineuse et les quelques succès obtenus l'ont surtout été dans la cognition de bas niveau (dans le contrôle moteur, par exemple).

La TDE rompt donc totalement avec la TCE et il semble bien impossible de tenter une réconciliation entre ces deux conceptions de la cognition. Toutefois, on pourrait aussi considérer que la TCE et la TDE ne décrivent pas les mêmes niveaux d'organisation de la cognition. C'est de cette façon qu'il faut comprendre les *modèles hybrides* ou *symbolicoconnexionnistes*. La TCE et la TDE décriraient plutôt, de façon non contradictoire mais complémentaire, des niveaux d'organisation différents : processus de haut niveau gouvernés par des représentations symboliques et des règles, pour la première, et processus de bas niveau, perceptivomoteurs, de type classification de formes ou contrôle du mouvement, pour la seconde.

La TRE postule la complétude et la relative permanence des états mentaux. Mais les représentations mentales sont loin d'être complètes (par exemple, le phénomène de cécité au changement) et elles sont souvent instables en fonction, en particulier, de la nature et de l'organisation des apprentissages perceptifs. Soutenir, de façon extrême, la TRE semble donc intenable. Mais défendre un relativisme absolu serait tout aussi facile à infirmer expérimentalement (par exemple, le phénomène de constance perceptive). Une solution hybride semble ainsi plus raisonnable et en accord avec les données empiriques : les représentations mentales du monde sont plus ou moins complètes et leur accessibilité, sinon leur disponibilité, n'est pas toujours garantie et dépend de la situation de l'agent cognitif dans un environnement unique car toujours renouvelé (Markman et Dietrich, 2000 ; Plagnol, ce numéro).

La TRE postule également que les possibilités d'adaptation de l'agent cognitif reposent sur un système amodal de représentations abstraites (par exemple : frames, scripts, prototypes, unités de reconnaissance faciale, etc.). Ce système se modifie donc sous l'influence d'un processus de généralisation à partir de l'expérience perceptive et selon un mécanisme de type descendant *top-down*. Sur le plan empirique, on devrait donc s'attendre à ce que la formation d'un concept ou la catégorisation d'un objet ne soit pas influencée par la nature et l'ordre des expériences perceptives qui en sont à l'origine. On sait que ce n'est pas le cas (Rodet et Tiberghien, 1994 ; Schyns et Rodet, 1997). Adopter le postulat opposé ne serait guère plus satisfaisant : si la flexibilité ne dépendait que de l'expérience perceptive, l'agent cognitif se modifierait sous l'influence d'un processus de différenciation sans fin. La seule issue est donc d'accepter un postulat hybride selon lequel les représentations mentales ne seraient pas constituées uniquement de symboles abstraits (amodaux) mais aussi de symboles perceptifs.

La TRE décrit le monde représenté sans le mettre en rapport avec les systèmes transducteurs, sensoriels et moteurs, qui assurent l'interface avec le monde réel. Cette cognition n'a donc pas vraiment de corps, c'est une cognition idéale. Elle n'appartient pas non plus à un

corps social, c'est une cognition individuelle. Poser le principe de l'inscription corporelle de la cognition c'est, au contraire, accorder une importance décisive aux interactions sensorimotrices avec l'environnement et aux interactions sensorimotrices entre de multiples agents cognitifs (sur cette question des rapports entre le corps et la cognition, voir : Andrieu, 2002, ce numéro ; Tête, 2002). Cette façon, très gibsonienne, de décrire la cognition est clairement illustrée par le concept d'affordance selon lequel les propriétés objectives des objets (sur-)déterminent nos perceptions, nos souvenirs et nos actions. Inscrire la cognition dans le corps c'est évidemment considérer que nos représentations mentales sont construites de façon ascendante *bottom-up*. Toutefois ces deux conceptions, portées à l'extrême, sont intenables si on les confronte, sans biais de sélection, aux données empiriques dont on dispose dans les sciences cognitives. Par exemple, la reconnaissance des visages, peut être facilitée ou perturbée par des propriétés sensorielles de bas niveau mais aussi par des représentations sémantiques ou épisodiques de haut niveau (Baudouin et Tiberghien, 2002). Ici encore, un postulat hybride, acceptant à la fois des processus ascendants et descendants, est sans doute le meilleur choix.

En somme, le seul point d'accord entre toutes ces théories, c'est que l'agent cognitif a des états internes ! C'est un accord vraiment minimaliste. De toute façon, d'un point de vue pratique, la TCE réussit bien mieux dans l'explication des processus cognitifs de haut niveau et la TDE dans celle des processus cognitifs de bas niveau. S'agit-il vraiment de deux conceptions antagonistes ou s'agit-il de théories décrivant des niveaux différents d'organisation de la cognition, ou plutôt du système « cerveau-esprit » ? La question est ouverte et certains philosophes y ont répondu positivement (Engel, 1991, 1994). Si tel était le cas, on peut alors espérer qu'une théorie hybride de la cognition associant, de façon interactive, un système dynamique d'adaptation cognitive (le cortex médiotemporal, par exemple ?) à un système représentationnel de stockage distribué (le néocortex frontal, temporal et pariétal, par exemple ?) et un système fonctionnel de récupération (l'hippocampe, par exemple ?) puisse rendre compte d'un grand nombre de données empiriques. Comme l'a reconnu récemment Fodor (2000–2003), « nous avons toutes les raisons de supposer que la théorie computationnelle constitue *une partie* de la vérité sur la cognition » (p. 9). Dès lors, il semble bien que seule une théorie hybride puisse expliquer à la fois la partie modulaire de l'architecture cognitive, mais aussi sa partie non modulaire, celle qui permet des inférences cognitives abductives, c'est-à-dire des inférences globales, sensibles aux systèmes de croyance, à la mémoire et aux variations contextuelles.

3. Neuro-imagerie cognitive : nouvel indicateur cognitif ou nouvelle phrénologie ?

Nous avons défini précédemment le projet des sciences cognitives comme une tentative d'explication de l'interface entre causalité mentale et causalité cérébrale. La neuro-imagerie offre un ensemble de techniques prometteuses qui peuvent aider à réaliser un tel programme. Mais elle peut aussi engendrer des illusions qui constitueront de sérieux obstacles épistémologiques aux progrès de nos connaissances dans ce domaine.

La neuro-imagerie ne devrait évidemment pas être considérée comme une discipline autonome, voire comme une nouvelle science. Plus modestement, il s'agit d'une technologie offrant de nouveaux indicateurs du comportement et de la cognition. Ce n'est pas non plus un indicateur universel devant se substituer aux autres indicateurs de la cognition ou en réduire l'importance. Si on peut la concevoir comme un microscope de la psychologie, il y en a d'autres et ils nous donnent à voir des objets différents à des échelles variables. Comme tout autre indicateur cognitif, il doit donc être interprété et le fait que des données quantitatives se

présentent sous la forme d'une image ne leur donne aucun caractère d'évidence supplémentaire. Son interprétation dépend d'ailleurs de fortes contraintes métrologiques et théoriques.

3.1. *Problèmes techniques et méthodologiques de la neuro-imagerie*

Une neuro-image, avant d'être une image, est évidemment une mesure, l'enregistrement d'un signal, et celui-ci suppose d'accepter certains postulats ainsi que des règles de définition, de construction et d'enregistrement précises. Par exemple, dans la tomographie par émission de positons (TEP), la mesure locale du flux sanguin dans le cerveau est corrélée à l'activité cellulaire, mais cette corrélation, par définition, n'est pas une déduction causale et elle est d'ailleurs encore inexpliquée. De plus, si le changement de débit sanguin s'accompagne d'une modification de la consommation d'oxygène, les changements métaboliques associés à l'activité neuronale, a fortiori cognitive, ne suivent pas précisément le couplage temporel entre débit sanguin et métabolisme d'oxydation du glucose. En d'autres termes, la corrélation entre les états cognitifs et les états du signal neuronal est une corrélation partielle, et leur synchronisation temporelle est loin d'être parfaite. On comprend, dans ces conditions, la complexité du signal enregistré car il résulte d'une cascade de corrélations entre des mécanismes physicochimiques dont la structure causale n'est pas encore complètement élucidée.

De plus, le signal enregistré, en PET ou en IRMf, est un simple signal d'activation qui ne peut en aucun cas être interprété en termes d'excitation ou d'inhibition. Cela a pour conséquence qu'un même pattern d'activation observé dans deux situations différentes peut avoir une signification fonctionnelle différente. Sans oublier aussi les effets résiduels des modifications hémodynamiques qui peuvent durer près d'une minute, les artefacts divers liés aux mouvements du sujet, à sa respiration et à son rythme cardiaque et, enfin, l'hétérogénéité du rapport signal-bruit selon les différentes régions du cerveau (Buckner et Logan, 2001).

Ces problèmes techniques se combinent, de plus, avec des difficultés méthodologiques qui sont loin d'être négligeables. On sait que l'analyse des données en neuro-imagerie repose sur la méthode soustractive. Cette méthode, classique en psychologie cognitive, a été utilisée avec succès pour dissocier des processus cognitifs, non observables, modifiant la latence des comportements dans des tâches dans lesquelles ils sont, ou non, impliqués (Donders, 1865–1969 ; Sternberg, 1969a, 1969b). Ces techniques chronométriques ont largement contribué à l'essor de la psychologie cognitive et, en particulier, pour l'étude de la mémoire. Elles reposaient toutefois sur un modèle sous-jacent postulant une combinaison additive de la latence partielle des différents sous-processus contribuant à une réponse comportementale. Un tel modèle, linéaire, rendait donc l'étude des interactions entre processus cognitifs non observables, sinon impossibles, du moins particulièrement difficile et indirecte.

On a ainsi rapidement considéré que la neuro-imagerie permettait de dépasser les limitations apparues dans les études de chronométrie mentale. En effet, en imagerie cérébrale, tout changement d'un processus entraîne un changement spatial observable directement dans le cerveau et le problème des interactions cachées est, en principe, éliminé. Comme l'a écrit Raichle (2001), « les événements qui surviennent dans le cerveau ne sont pas cachés à l'investigateur comme ils le sont dans les expériences purement cognitives » (pp. 8–9). Certes, mais les entités cognitives le demeurent, et la relation entre les événements cérébraux et ces dernières ne peut se lire de façon directe que si l'on continue à accepter le postulat d'un modèle sous-jacent de type linéaire (additif).

On comprend, dès lors, que la stratégie de localisation cérébrale des processus cognitifs ait prévalu : le meilleur moyen de ne pas avoir de problèmes d'interaction est de les faire dispa-

raître, au sens propre de ne pas les voir dans les neuro-images². Mais un changement d'activité dans une région définie du cerveau pourrait bien être associé à des opérations cognitives différentes s'il se manifestait dans des patterns d'activation impliquant d'autres régions cérébrales. En d'autres termes, la méthode soustractive en neuro-imagerie permet d'isoler une ou plusieurs régions cérébrales critiques dont la différence d'activation, entre une condition témoin et une condition expérimentale, est la plus élevée, et cela en fonction d'un certain critère statistique (souvent un simple test *z*). Cet effet est alors attribué à une entité cognitive hypothétique que l'on postule unitaire, c'est-à-dire non-décomposable en d'autres sous-processus cognitifs, de plus bas niveau, ou non-attribuables à des processus cognitifs de plus haut niveau. C'est un pari évidemment à très haut risque et qui peut même être théoriquement et logiquement contesté, car il repose à la fois sur une conception strictement modulariste de la cognition et sur un modèle linéaire du cerveau.

Ainsi, localiser une entité cognitive dans une région cérébrale n'a de sens (cognitif) que si l'on présente le pattern le plus complet possible des intercorrélations entre l'activation de cette région et celles des autres régions du cerveau. Ce pattern fait justement apparaître les interactions occultées par la méthode soustractive standard. Un tel pattern d'interactions peut alors être analysé au moyen des méthodes d'analyse statistique multidimensionnelles (pour une application de ces méthodes, McIntosh, 1999 ; Cabeza, 2001). Par exemple, au lieu de considérer que la mémoire de travail est localisée dans le cortex préfrontal, ce dernier type de méthodes aboutit à une description beaucoup plus complexe de la réalité selon laquelle la mémoire de travail est un ensemble de propriétés émergentes des interactions entre le cortex préfrontal et d'autres régions du cerveau.

Les difficultés précédentes sont probablement mieux prises en compte par les techniques d'enregistrement des potentiels évoqués liés à processus cognitifs. En effet, ces dernières décrivent précisément la dynamique temporelle de la réponse. Par exemple, dans les 600 ms qui séparent la présentation d'un visage de la réponse de reconnaissance, on constate qu'un grand nombre de régions cérébrales différentes sont activées, de façon antérograde, bien sûr, mais aussi de façon rétrograde.

Le problème de la localisation cérébrale des sous-processus cognitifs de familiarité (F) et de récollection (R), qui sont supposés intervenir dans la reconnaissance mnésique, est particulièrement instructif à cet égard. Les modèles neuroanatomiques récents, ayant utilisé la TEP ou l'IRMf, ont suggéré que l'hippocampe était critique pour les reconnaissances basées sur R et que les structures parahippocampiques étaient critiques pour les reconnaissances basées sur F (Yonelinas, 2002). Toutefois, dans une étude de reconnaissance de visages, avec la méthode des potentiels évoqués, nous avons montré que R et F n'avaient pas nécessairement des substrats neurophysiologiques différents, mais impliquaient plutôt une énergie et un décours temporel différents à l'intérieur de réseaux neuronaux fonctionnels distincts (Guillaume et Tiberghien, 2002, 2005).

La mise en évidence de covariances cérébrales antéropostérieures s'accorde bien, d'ailleurs, avec l'idée que les processus de reconnaissance opèrent en cascade et impliquent de nombreuses rétroactions. C'est donc bien le postulat d'indépendance fonctionnelle de la familiarité et de la récollection qui est ici remis en cause. La méthode des potentiels évoqués permet en

² De la même façon, dans les études comportementales, le meilleur moyen de faire disparaître les interactions est de substituer une série d'expériences indépendantes (à un seul facteur) à une expérience unique prenant en compte, simultanément, l'ensemble de ces facteurs. Cette pratique expérimentale, très répandue, est en fait une dissimulation, plus ou moins consciente, de la complexité.

effet d'analyser le décours temporel des sous-processus mis en jeu en reconnaissance : d'abord la récupération de l'image du visage sous sa forme globale (familiarité) puis la récupération des détails perceptifs (récollection). Si l'on décompose ces deux moments sur les sites antérieurs et postérieurs du cerveau, des différences fonctionnelles apparaissent clairement :

- la première période correspond à l'initialisation et à l'orientation de la reconnaissance sur les informations nécessaires au jugement de reconnaissance faciale (cortex frontal) et, dans le même temps, à la récupération globale, plus ou moins efficace, de la trace mnésique du visage (cortex temporopariétal) ;
- la seconde période correspond plutôt à la récupération contrôlée et dirigée de façon explicite sur des informations faciales particulières (au niveau temporopariétal), au contrôle et à la validation de cette récupération en fonction de la situation de reconnaissance et du contexte de traitement (au niveau frontal).

3.2. *Problèmes théoriques de la neuro-imagerie*

Cet ensemble de difficultés techniques et méthodologiques aurait dû conduire à une relative prudence dans les interprétations théoriques des localisations cérébrales. Malheureusement, c'est loin d'être la règle. Il est pourtant clair que l'on ne peut observer dans le cerveau que les entités cognitives qui ont été construites par d'autres méthodes (expérimentation comportementale, modélisation) et en utilisant d'autres indicateurs. En d'autres termes, l'interprétation d'une image cérébrale dépend des théories et modèles cognitifs dont on dispose. Bien évidemment, « on ne voit pas le cerveau penser » (Decety, 1999), ni des « images de l'esprit » (Posner, 1994). Ce ne sont là que des métaphores. Au contraire, les chercheurs projettent en permanence leurs modèles cognitifs dans l'activité métabolique du cerveau. Ce sont d'ailleurs ces modèles que l'on voit dans le cerveau ou, plutôt, ce sont ces modèles qui nous permettent de donner du sens à ce que l'on voit dans le cerveau et, comme le rappelle Le Ny (2005), « le sens n'est pas un observable » (p. 16). Ainsi, de ce point de vue, et selon la belle expression de Arbib (1989), le cerveau est un objet métaphorique.

Une conséquence inattendue et provocatrice de ce paradoxe pourrait être la suivante : la neuro-imagerie ne permettra sans doute jamais d'apporter des connaissances théoriques nouvelles dans le domaine d'étude de la cognition³. Mais elle permettra sans doute, plus modeste-

³ J'ai posé récemment à un neurophysiologiste la question suivante : « Quelle connaissance nouvelle la neuro-imagerie a-t-elle apportée sur le fonctionnement cognitif ? La seule découverte qu'il ait mentionnée (mais le jeu est ouvert !) a été celle des neurones-miroirs : neurones qui sont activés quand on effectue une action, mais aussi quand on l'anticipe, ou qu'on l'observe chez autrui (Gallese et Goldman, 1998 ; Jeannerod, 2002 ; P. Jacob, ce numéro). Son argumentation était la suivante : l'état d'observation ou d'anticipation est un état mental non associé à un comportement et il ne peut donc être observé que par l'intermédiaire d'un indicateur cérébral observable, dans des conditions optimales, par des techniques de neuro-imagerie. Sa surprise a été grande quand je lui ai dit qu'en 1929, Blodgett (cité par Tolman, 1948) avait montré qu'un groupe de rats, observant d'autres rats témoins engagés dans un apprentissage de labyrinthe, bénéficiait de cette observation et atteignait ensuite le critère d'apprentissage plus rapidement que les animaux témoins. Ce phénomène, connu sous le nom d'apprentissage latent, a donné lieu, à l'époque, à de nombreuses recherches. Tolman a d'ailleurs invoqué ce phénomène, et celui des essais et erreurs vicariants *vicarious trial and error* ou VTE), pour fonder sa théorie des cartes mentales et de l'intention (*purpose*). Ainsi, la mise en évidence de neurones-miroir par la neuro-imagerie confirme un phénomène qui avait déjà été mis en évidence avec les seules méthodes comportementales. L'apport méthodologique de la neuro-imagerie a donc été ici de montrer que l'activité mentale (observation, anticipation), inférée à partir de la performance d'apprentissage, correspondait bien à une modification objective au niveau cérébral.

ment mais très efficacement, de mettre en évidence les réseaux neuronaux qui participent à certaines opérations cognitives et, par conséquent, de valider les modèles qui les ont inventées à partir d'approches multiples. Que serait la neuro-imagerie cognitive sans la psychologie du comportement et sans la psychologie cognitive ?

Une fonction cognitive ne peut donc être localisée de façon absolue dans le cerveau, car comme nous avons essayé de le montrer précédemment, ses différents composants font intervenir, à différents moments, des ensembles neuronaux variés et largement distribués. Je ne prendrai ici qu'un exemple de la complexité des inférences cognitives à partir des données d'imagerie cérébrale. Un grand nombre de recherches, conduites à la fois sur l'homme et sur l'animal, au moyen de diverses techniques d'enregistrement de l'activité cérébrale, ont abouti à la conclusion d'une localisation du traitement de l'information faciale dans le lobe temporal (gyrus fusiforme latéral) [Gross et Sergent, 1992]. Mais d'autres recherches (Farah et al., 1998 ; Gauthier et al., 2000) ont aussi montré que cette région était activée dans la reconnaissance experte d'objets différents des visages (voitures, bâtiments, animaux, etc.). Il est donc possible de considérer que cette région est associée, non au traitement de l'information faciale, mais à l'expertise visuelle en général, sa sensibilité aux visages n'étant que la conséquence du très haut niveau d'expertise faciale communément observé dans l'espèce humaine et chez les primates non humains. On peut donc localiser, au même endroit dans le cerveau, un processus cognitif (la reconnaissance des visages) ou un processus de plus haut niveau auquel le premier est subordonné (l'expertise visuelle). L'inverse est également possible. Par exemple, le traitement de l'expression a été localisé dans le sulcus temporal supérieur (STS) par Hasselmo et al. (1989), mais Allison et al. (2000) y localisent la perception sociale (processus de plus haut niveau se basant, en partie, sur l'expression), tandis que Haxby et al. (2000) y localisent le traitement des aspects variables du visage (processus de plus bas niveau nécessaire au traitement de l'expression) et, enfin, Desimone (1997) y localise le traitement focalisé ou local (processus de plus bas niveau nécessaire au traitement des aspects variables du visage) (pour une discussion de cette question : Tiberghien et al., 2003).

Ces incertitudes ont une cause théorique évidente. On ne peut voir dans le cerveau que ce que l'on y apporte et le manque de consensus, au niveau de la description des entités cognitives, ne peut être réglé par la seule précision des localisations cérébrales, bien au contraire. On peut certes admettre qu'il soit possible de localiser avec une relative précision les corrélats neuronaux d'un comportement en réponse à une situation, à condition que ceux-ci soient décrits à un niveau de précision très élevé, je veux dire à un niveau de précision psychophysique, atteint généralement dans les recherches sensorimotrices. Mais cet objectif est sans doute encore difficilement réalisable quand il s'agit de localiser très précisément des processus cognitifs dont la définition théorique ou opérationnelle fait encore débat. D'une part, les entités cognitives invoquées sont souvent définies de façon imprécise et leurs indicateurs comportementaux prêtent eux-mêmes à discussion, d'autre part, ils donnent lieu à une décomposition théorique en sous-processus cognitifs qui est rarement consensuelle.

Par exemple, comment définir le concept cognitif de familiarité ? Ce concept a-t-il le même sens dans les situations perceptives, les situations mnésiques et les situations de raisonnement ? Comment se différencie-t-il, précisément, du concept de fluence perceptive ? Comment mesurer la familiarité : par un jugement, la vitesse de la réponse, son intensité, sa certitude ? Quelle confiance métrologique faut-il accorder aux évaluations verbales de familiarité ? Quelle est la relation entre les mesures de la familiarité ? Etc. Inutile de préciser qu'il n'y a pas de consensus sur ces questions et une grande variété de théories et de modèles est ici en compétition.

Ou encore, autre exemple, comment faut-il comprendre des assertions du genre : le regret « impliquerait » le cortex orbitofrontal (Camille et al., 2004) ? Peut-être, mais peut-être que d'autres régions cérébrales sont aussi impliquées à des degrés divers dans cet état cognitif. De toute façon, il est douteux qu'il y ait accord sur la définition théorique et sur la mesure d'une entité cognitive aussi complexe ? De plus, il est vraisemblable qu'un ensemble de sous-processus en interaction constante détermine un tel état de conscience : mémoire, raisonnement, attention, émotion, expectation, motivation, estimation subjective des probabilités, estimation subjective des gains et des coûts, décision, par exemple. Par ailleurs, la nature et l'intensité des situations engendrant regret et déception (est-ce, au fait, la même chose ?) sont des facteurs qui doivent largement moduler la combinaison résultante de ces sous-processus.

3.3. Nouvelle phrénologie

Doit-on, dans ces conditions, accepter de façon non critique des données d'imagerie cérébrale qui, avec une précision parfois extrême, localisent dans le cerveau des entités cognitives dont la définition théorique et les indicateurs comportementaux demeurent problématiques ? Évidemment non ! Si tel était notre choix méthodologique, il y aurait un risque élevé de laisser se développer une nouvelle phrénologie, plus pernicieuse encore que la première, car semblant confirmée par les technologies les plus modernes (sur la définition et l'histoire de la première phrénologie : Lanteri-Laura, 1970 ; Tiberghien et al., 2002 ; Nicolas, ce numéro. Sur la nouvelle phrénologie : Uttal, 2001).

La neurologie ayant montré que les fonctions sensorielles et motrices étaient relativement bien localisées dans le cerveau, il était évidemment séduisant de prédire que les fonctions cognitives le seraient également. Mais, nous avons vu que celles-ci sont souvent définies de façon approximative et problématique. Elles sont souvent d'ailleurs reliées de façon lâche aux indicateurs comportementaux et aux entrées sensorielles. Par quel miracle des entités cognitives, mal ou peu définies, pourraient-elles être localisées précisément dans le cerveau ? Que penser de ces recherches, publiées dans des revues apparemment sérieuses, qui trouvent dans le cerveau aussi bien les neurones de la sagesse que les neurones de la sympathie. On a même trouvé une aire cérébrale de la transe mystique, une aire du nirvana et, il fallait s'y attendre, un module de Dieu car « le cerveau humain a été génétiquement conçu pour encourager les croyances religieuses » (Newberg et al., 2001) ! Quel est l'intérêt, scientifique, méthodologique ou théorique, à localiser les régions activées dans la méditation (Lou et al., 1999), l'expérience religieuse (Azari et al., 2001), le jugement empathique et le pardon (Farrow et al., 2001) ou l'amour romantique (Bartels et Zekin, 2000) ? À mon sens, il n'y en a pas, et ces recherches utilisent simplement le prestige technologique de la neuro-imagerie afin de justifier des idées reçues du sens commun ou des conceptions psychologiques, sociales, et même politiques, largement dominantes.

Il n'est donc pas très surprenant de constater que des recherches de plus en plus nombreuses tentent de fonder scientifiquement la manipulation des esprits sur des techniques de neuro-imagerie. Des laboratoires de recherche fondamentale⁴ vendent à des entreprises des promes-

⁴ Citons, aux États-Unis le laboratoire Mind and the Market (Harvard Business School) patronné par un psychologue (S.M. Kosslyn) et un économiste (G. Zaltman). En France, un psychologue (O. Koenig), un neurologue (B. Croisille) et un publicitaire (B. Poyet) ont créé, en 2001, un cabinet de consultants « *Impact Mémoire* » dont l'objectif est d'utiliser nos connaissances, psychologiques et neurologiques, dans le domaine de la mémoire afin d'optimiser les messages publicitaires. C'est du très vieux vin dans de nouvelles bouteilles. Bien que nous ayons un avis très précis sur ce genre d'entreprises, nous préférons laisser le lecteur s'en faire lui-même une opinion précise, en consultant, par exemple, les sites Internet concernés (Oullier, 2003 ; Benoît-Browaecs, 2005).

ses, sans doute illusoirs, d'images cérébrales associées à des produits commerciaux. On parle maintenant de neuroéconomie (libérale, bien sûr), de neuromarketing (idem), de neuroéthique (ibidem), et on identifie, sans gêne, la « zone de la récompense », la « zone cérébrale du marché » ou, encore plus fort, les « zones de préférence du Coca-Cola » (Amalou, 2001 ; Montague et McClure, 2004). Ce n'est, hélas, pas la première fois que la psycho- et la neurotechnologie sont mises au service d'un contrôle des esprits à des fins mercantiles ou politiques (Schwitzgebel et Schwitzgebel, 1973 ; Tiberghien, 1979). Cette technophrénologie ressemble d'ailleurs, à s'y méprendre, à l'ancienne et on lui souhaite d'ailleurs la même fin : qui se souvient, aujourd'hui, des céphalomètres, craniomètres, cranioscopes, goniomètres, stéréographes, physionotypes, profilomètres, anthropomètres et de tous les appareils pour le cubage des crânes qui faisaient la fierté du laboratoire d'anthropologie de l'École des Hautes Études, à Paris, dans les années 1870 ?

4. Conclusion

Nous avons voulu ici défendre et illustrer trois idées qui sont en fait très fortement liées :

- vouloir réduire les sciences cognitives à l'une de ses composantes, c'est en fait détruire le programme même des sciences cognitives. Celles-ci ne sont ni des neurosciences cognitives, ni des psychosciences cognitives, ni des philosciences cognitives ;
- une théorie représentationnelle de l'esprit ne permet pas de rendre compte de la totalité de l'activité cognitive et elle est, à tout le moins incomplète. Il en est de même d'ailleurs pour une théorie dynamique de l'esprit (connexionnisme radical). Une théorie hybride de l'esprit est inévitable et semble seule capable d'intégrer, pour le moment, l'ensemble des données empiriques disponibles ;
- enfin, les méthodes de neuro-imagerie offrent aux sciences cognitives un nouveau moyen d'étude de la cognition, l'indicateur cérébral qui lui manquait. Mais celui-ci n'a de sens que s'il est mis en relation de façon non-ambiguë avec les autres indicateurs de la cognition. La surestimation de sa signification théorique et de sa valeur méthodologique conduit inévitablement à localiser, de façon compulsive, des entités cognitives (souvent mal définies), et qui mettent sans doute en œuvre une interaction complexe entre de nombreux systèmes cérébraux.

À la conférence Hixon (1948), Lashley a déclaré : « Le cerveau est un système dynamique constamment actif *ou, plutôt*, un ensemble de nombreux systèmes en interaction ». Pour l'explication des activités mentales, en particulier les plus complexes, il avait fondamentalement raison. Une telle position ne peut être réduite, sinon de façon caricaturale, à une simple théorie de l'équipotentialité de toutes les régions du cerveau. Plus un processus cognitif est complexe, et plus il fait intervenir, dans l'espace et le temps cérébral, un nombre important de sous-systèmes selon une dynamique non linéaire. Vouloir localiser un tel processus dans une région restreinte du cerveau, est un vieux rêve (le rêve de F.J. Gall), dont il faudrait sortir le plus rapidement possible.

Références

- Allison, T., Puce, A., McCarthy, G., 2000. Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences* 4, 267–278.

- Amalou, F., 2001. Fouiller la mémoire pour accroître l'effet des publicités. *Le Monde*.
- Andrieu, B., 2002. *La chair du cerveau. Phénoménologie et biologie de la cognition*. Sils Maria, Mons (Belgique).
- Arbib, M.A., 1989. *The metaphorical brain 2. Neural networks and beyond*. Wiley, New York.
- Azari, N.P., Nickel, J., Wunderlich, G., Nieddeggen, M., Hefter, H., Tellmann, L., et al., 2001. Neural correlates of religious experience. *European Journal of Neuroscience* 13, 1649–1652.
- Bachelard, G., 1965. *La formation de l'esprit scientifique. Contribution à une psychanalyse de la connaissance objective*. Librairie Philosophique J. Vrin, Paris.
- Bartels, A., Zekin, S., 2000. The neural basis of romantic love. *NeuroReport* 11, 3829–3834.
- Benoît-Browaeys, D., 2005. Jusqu'où ira le neuromarketing ? *Alternatives Économiques* N° 232.
- Baudouin, J.-Y., Tiberghien, G., 2002. Gender is a dimension of face recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28, 362–365.
- Bohr, N., 1932/1959/1961. *Physique atomique et connaissance humaine*. Gonthier, Paris.
- Bonnet, C., 2000. Psychologie du comportement et psychologie cognitive. *Psychologie Française* 45, 3–11.
- Buckner, R.L., Logan, J.M., 2001. Functional neuroimaging methods: PET and fMRI. In: Cabeza, R., Kingstone, A. (Eds.), *Handbook of functional neuroimaging of cognition*. The MIT Press, Cambridge, MA, pp. 27–48.
- Cabeza, R., 2001. Functional neuroimaging of cognitive aging. In: Cabeza, R., Kingstone, A. (Eds.), *Handbook of functional neuroimaging of cognition*. The MIT Press, Cambridge, MA, pp. 331–377.
- Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J.-R., Sirigu, A., 2004. The involvement of orbito-frontal cortex in the experience of regret. *Science* 304, 1167–1170.
- Canguilhem, G., 1958. Qu'est-ce que la psychologie ? *Revue de Métaphysique et de Morale* 1, 12–25.
- Churchland, P.M., 1986. *Neurophilosophy: Toward a unified science of mind-brain*. The MIT Press, Cambridge, MA.
- Decety, J., 1999. Voir le cerveau penser : intérêt et limites des techniques de neuro-imagerie. *Annales Médico-Psychologiques* 157, 673–686.
- Delay, J., 1945–1963. *La psychophysiologie humaine*. Presses Universitaires de France, Paris.
- Desimone, R., 1997. Attention control of visual perception. *Electroencephalography and Clinical Neurophysiology* 102, 4.
- Donders, F.C., 1865–1969. On the speed of mental process. *Acta Psychologica* 30, 412–431.
- Dupéron, I., 2000. *G.T. Fechner : le parallélisme psychophysique*. Presses Universitaires de France, Paris.
- Dupuy, J.-P., 1994. *Aux origines des sciences cognitives*. La Découverte, Paris.
- Engel, P., 1991. Psychologie populaire et explication cognitive. In: Missa, J.N. (Ed.), *Philosophie de l'esprit et sciences du cerveau*. Librairie Philosophique J. Vrin, Paris, pp. 135–146.
- Engel, P., 1994. *Introduction à la philosophie de l'esprit*. La Découverte, Paris.
- Farah, M.J., Wilson, K.D., Drain, M., Tanaka, J.N., 1998. What is "special" about face perception? *Psychological Review* 105, 482–498.
- Farrow, T.F., Zheng, Y., Wilkinson, I.D., Spence, S.A., Deakin, J.F., Tarrier, N., et al., 2001. Investigating the functional anatomy of empathy and forgiveness. *NeuroReport* 12, 2433–2438.
- Fodor, J.A., 2000–2003. *L'esprit, ça ne marche pas comme ça : portée et limites de la psychologie computationnelle*. Éditions Odile Jacob, Paris.
- Fraisse, P., Piaget, J., 1966. *Traité de psychologie expérimentale (Vol. III. Psychophysiologie du comportement)*. Presses Universitaires de France, Paris.
- Gallese, V., Goldman, A., 1998. Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences* 2, 493–501.
- Gardner, H., 1985–1993. *Histoire de la révolution cognitive : la nouvelle science de l'esprit*. Payot, Paris.
- Gauthier, I., Skudlarski, P., Gore, J.C., Anderson, A.W., 2000. Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience* 3, 191–197.
- Guillaume, F., Tiberghien, G., 2002. Conscious memory retrieval and decision processes reveal by the P300 theta oscillation during face recognition. *Brain and Cognition* 53, 110–116.
- Guillaume, F., Tiberghien, G., 2005. An electrophysiological study of contextual variations in a short-term face recognition memory. *Cognitive Brain Research* 22, 471–487.
- Gross, C.G., Sergent, J., 1992. Face recognition. *Current Opinion in Neurobiology* 2, 156–161.
- 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.
- Haxby, J.W., Hoffman, E.A., Gobbini, M.I., 2000. The distributed human neural system for face perception. *Trends in Cognitive Sciences* 4, 223–233.
- Jacob, P., 2000. Philosophie de l'esprit et sciences cognitives. In: Michaud, Y. (Ed.), *Université de tous les savoirs (Vol. 3). Qu'est-ce que la société ?* Éditions Odile Jacob, Paris, pp. 19–28.
- James, W., 1890/1891. *The principles of psychology (Vol. 1)*. MacMillan, London.

- Jeannerod, M., 1996. *De la physiologie mentale : histoire des relations entre biologie et psychologie*. Odile Jacob, Paris.
- Jeannerod, M., 1997. Vers un darwinisme mental ? La pensée évolutionniste en neurosciences. In: Tort, P. (Ed.), *Pour Darwin*. Presses Universitaires de France, Paris, pp. 288–307.
- Jeannerod, M., 2002. *La nature de l'esprit*. Éditions Odile Jacob, Paris.
- Jouvet, M., 2004. *Le voleur de songes*. Éditions Odile Jacob, Paris.
- Kosslyn, S.M., Koenig, O., 1992. *Wet mind: The new cognitive neuroscience*. The Free Press, New York.
- Lanteri-Laura, G., 1970. *Histoire de la phrénologie l'homme et son cerveau selon F.J. Gall*. Presses Universitaires de France, Paris.
- Le Ny, J.-F., 2005. *Comment l'esprit produit du sens*. Editions Odile Jacob, Paris.
- Lou, H.C., Kjaer, T.W., Friberg, L., Wildschioodtz, G., Holm, S., Nowak, M., 1999. A 150-H2O PET study of meditation and the resting state of normal consciousness. *Human Brain Mapping* 7, 98–105.
- Markman, A.B., Dietrich, E., 2000. Extending the classical view of representation. *Trends in Cognitive Sciences* 4, 470–475.
- McIntosh, A.R., 1999. Mapping cognition to the brain through neural interactions. *Memory* 7, 523–548.
- Montague, R., McClure, S., 2004. Neural correlates of behavioral preference for culturally familiar drinks. *Neuron* 44, 379–387.
- Netchine-Grynberg, G., Netchine, S., 1989. À propos de quelques fictions dans la référence au biologique en psychologie : réduction ou formation de champs intermédiaires. In: Mengal, P., Parot, F. (Eds.), *La fabrique, la figure et la feinte : fictions et statut des fictions en psychologie*. Librairie Philosophique J. Vrin, Paris, pp. 81–101.
- Newberg, A., D'Aquilli, E., Rause, V., 2001. *Why God won't go away. Brain science and the biology of belief*. Ballantine Books, New York.
- Oullier, O., 2003. Le « neuromarketing » est-il l'avenir de la publicité ? *Le Monde*, 17.
- Pacherie, E., 1993. *Naturaliser l'intentionnalité : essai de philosophie de la psychologie*. Presses Universitaires de France, Paris.
- Pacherie, E., Proust, J. (Eds.), 2004. *La philosophie cognitive*. Fondation de la Maison des sciences de l'homme. Éditions Ophrys, Paris.
- Piaget, J., 1963–1967. L'explication en psychologie et le parallélisme psychophysique. In: Fraisse, P., Piaget, J. (Eds.), *Traité de psychologie expérimentale* (Vol. 1). Histoire et méthode. Presses Universitaires de France, Paris, pp. 123–162.
- Piaget, J., 1965. *Sagesse et illusions de la philosophie*. Presses Universitaires de France, Paris.
- Piaget, J., 1970. *Psychologie et épistémologie : pour une théorie de la connaissance*. Denoël, Paris.
- Posner, M.I., 1994. *Images of mind*. Freeman, New York.
- Putnam, H., 1973. Reductionism and the nature of psychology. *Cognition* 2, 131–146.
- Putnam, H., 1988–1990. *Représentation et réalité*. Gallimard, Paris.
- Raichle, M.E., 2001. Functional neuroimaging: A historical and physiological perspective. In: Cabeza, R., Kingstone, A. (Eds.), *Handbook of functional neuroimaging of cognition*. The MIT Press, Cambridge, MA, pp. 3–26.
- Reuchlin, M., 1962. *Les méthodes quantitatives en psychologie*. Presses Universitaires de France, Paris.
- Reuchlin, M., 1965. Psychologie concrète et psychologie abstraite. *Psychologie Française* 10, 289–308.
- Richard, J.-F., Tiberghien, G., 1999. Épistémologie et psychologie : introduction. *Psychologie Française* 44, 193–196.
- Richelle, M., Seron, X., 1994. Le champ de la psychologie expérimentale : aspects historiques et épistémologiques. In: Richelle, M., Requin, J., Robert, M. (Eds.), *Traité de psychologie expérimentale* (Vol. 1). Presses Universitaires de France, Paris, pp. 3–42.
- Rodet, L., Tiberghien, G., 1994. Towards a dynamic model of associative semantic memory. *Journal of Biological Systems* 2, 401–411.
- Roy, J.-M., 2001. L'émergence de la neuroscience cognitive. *Cahiers Alfred Binet* 667, 9–33.
- Schwitzgebel, R.L., Schwitzgebel, R.K., 1973. *Psychotechnology: electronic control of mind and behavior*. Holt, Rinehart and Winston, New York.
- Schyns, P.G., Rodet, L., 1997. Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23, 681–696.
- Sternberg, S., 1969a. The discovery of processing stages: extensions of Donders' methods. *Acta Psychologica* 2, 276–315.
- Sternberg, S., 1969b. Memory scanning: mental processes revealed by reaction time experiments. *American Scientist* 57, 421–457.
- Tête, A., 2002. *La psychologie et ses fantômes*. Presses Universitaires de France, Paris.
- Tiberghien, G., 1979. Psychologie, idéologie et répression politique. *Psychologie Française* 24, 169–184.

- Tiberghien, G., 1985a. Fragments d'histoire de la psychologie. In: Mathieu, J., Thomas, R. (Eds.), *Manuel de psychologie*. Vigot, Paris, pp. 19–37.
- Tiberghien, G., 1985b. Mais où sont les stimulus d'antan ? *Psychologie Française* 30 (2), 177–183.
- Tiberghien, G., 1999. La psychologie cognitive survivra-t-elle aux sciences cognitives. *Psychologie Française* 44, 265–283.
- Tiberghien, G., Jeannerod, M., 1995. Pour la science cognitive : la métaphore cognitive est-elle scientifiquement fondée ? *Revue Internationale de Psychopathologie* 18, 173–203.
- Tiberghien, G., Abdi, H., Desclés, J.-P., Georgieff, N., Jeannerod, M., Le Ny, J.-F., Livet, P., Pynte, J., Sabah, G. (Eds.), 2002, *Dictionnaire des sciences cognitives*. Armand Colin, Paris.
- Tiberghien, G., Baudouin, J.-Y., Guillaume, F., Montoute, T., 2003. Should the temporal cortex be chopped in two? *Cortex* 39, 121–126.
- Tolman, E.C., 1948. Cognitive maps in rats and men. *Psychological Review* 55, 189–208.
- Uttal, W.R., 2001. *The new phrenology: the limits of localizing cognitive processes in the brain*. The MIT Press, Cambridge, MA.
- Varela, F.J., 1996. *Invitation aux sciences cognitives*. Seuil, Paris.
- Vauclair, J., Perret, P., 2003. The cognitive revolution in Europe: taking the developmental perspective seriously. *Trends in Cognitive Sciences* 7, 284–285.
- Woodworth, R.S., 1922/1930. *Psychology: a study of mental life*. Methuen, London.
- Yonelinas, A.P., 2002. The nature of recollection and familiarity: a review of 30 years of research. *Journal of Memory and Language* 46, 441–517.