

## **JALONS**

Jean-Pierre DUPUY. — *Sur le fonctionnalisme neumannien*

La notion de fonctionnalisme est aussi encombrante et mal dégrossie en philosophie de l'esprit et en sciences cognitives qu'elle l'a été en philosophie sociale et en sciences sociales. Le simple fait qu'elle prétende combiner le sens logico-mathématique et le sens téléologique du terme "fonction" dit assez, je crois, les confusions dont elle est grosse. Ma recommandation serait tout simplement que l'on décide, une fois pour toutes, de s'en passer. Mais enfin, puisqu'il faut jouer le jeu...

Soit une organisation, mentale ou sociale, que l'on peut décrire à différents "niveaux". Si, à un certain niveau, on peut établir un diagramme de relations causales entre les entités de ce niveau, et qu'il en résulte une explication acceptable du comportement global de l'organisation en question, on peut dire — au prix d'un laxisme répréhensible — que les entités du niveau "sont fonction" les unes des autres et que chacune a une "fonction" dans le comportement d'ensemble. Le fonctionnalisme entend en fait rendre légitimes, scientifiquement (c'est-à-dire causalement) parlant, des explications qui flirtent avec ces deux ennemis de la science galiléo-newtonienne: le holisme et le finalisme. Mais c'est encore un autre trait du fonctionnalisme qui rend compte des succès qu'il a obtenus, aussi bien en sciences sociales (par exemple, l'anthropologie d'un Malinowski ou la sociologie "structuralo-fonctionnaliste" d'un Parsons) qu'en sciences cognitives (par exemple le fonctionnalisme "computationnel-représentationnel" d'un Fodor ou d'un Pylyshyn). Dès lors qu'on dispose d'une explication valable du fonctionnement de l'organisation avec les seules ressources d'un niveau donné, l'indifférence à la façon dont les entités de ce niveau, ainsi que leurs relations causales, sont engendrées par, ou implémentées/réalisées dans, le niveau inférieur, se trouve en quelque sorte légitimée. Le fonctionnalisme peut donc se targuer de rendre compatibles un monisme ontologique et l'"autonomie relative" des niveaux d'organisation, et de fonder par là même une approche scientifique non réductionniste.

On ne reviendra pas ici sur la façon dont la révolution conceptuelle introduite par la machine de Turing a pu servir de base à une solution fonctionnaliste au problème des rapports entre la matière et la pensée, ni

sur toutes les distinctions afférentes (*token-physicalism/type-physicalism*, *supervenience*, monisme des particuliers/dualisme des propriétés, etc.), sinon pour dire ceci. Le problème de ce "monisme" matérialiste non réductionniste, c'est qu'il a une fâcheuse tendance à verser dans le dualisme ontologique. C'est ce que souligne Searle en écrivant, certes de façon polémique: "L'Intelligence Artificielle dans sa version forte ne prend sens que par rapport à l'hypothèse dualiste selon laquelle, pour qui s'intéresse à l'esprit, le cerveau n'a pas d'importance. En Intelligence Artificielle, aussi bien que pour le fonctionnalisme, ce qui compte, ce sont les programmes, et les programmes sont indépendants de leur réalisation dans des machines (...) C'est seulement si l'on croit que l'esprit est séparable du cerveau à la fois conceptuellement et empiriquement — c'est-à-dire si l'on est dualiste dans un sens fort — que l'on peut former l'espoir de reproduire le mental en écrivant et en faisant tourner des programmes, puisque les programmes doivent être indépendants des cerveaux ou de toute autre forme d'instantiation"<sup>1</sup>.

Il existe, en philosophie analytique de l'esprit, des tentatives de fonder un monisme non réductionniste autres que le fonctionnalisme turingien. Il n'est pas certain qu'elles réussissent mieux que ce dernier à éviter le piège dualiste. Je ne citerai que le "monisme anomal" de Davidson — lequel, il est intéressant de le souligner, trouve sa source d'inspiration dans le théorème de Tarski sur la vérité: le mental serait au physique ce que la sémantique est à la syntaxe. N'est-il pas significatif que Quine, qui, en dépit d'une philosophie personnelle proche du matérialisme éliminationniste, soutient le monisme anomal de Davidson, puisse écrire à son sujet: "L'ancien dualisme de l'esprit et du corps peut donc subsister, transmué et transplanté en un dualisme des concepts ou du langage. Comme tel, il demeure irréductible".<sup>2</sup>

Or il existe un moyen de se sortir de ce mauvais pas, c'est-à-dire de fonder un monisme non réductionniste qui ne verse pas dans le dualisme. Nous le devons au génie de John von Neumann, lors d'un moment fondateur de ce qui ne s'appelait pas encore les sciences cognitives. Je fais référence à sa conjecture sur la "complexité", formulée lors du symposium Hixon de 1948. Que les sciences cognitives soient restées, pour l'essentiel, aveugles aux potentialités de cette notion reste une énigme historique, pour laquelle je proposerai une ébauche de solution en conclusion.

---

<sup>1</sup> J. Searle, "Minds, Brains, and Programs", *The Behavioral and Brain Sciences*, tome 3, 1980.

<sup>2</sup> W.V.O. Quine, article "Mind versus Body", *Philosophical Dictionary*, 1987.

C'est dans ses travaux sur la théorie des automates que von Neumann devait préciser cette notion de complexité. En faisant une grandeur de type thermodynamique, il conjecturait qu'en-deçà d'un certain seuil, elle est dégénérative, le degré d'organisation ne pouvant que décroître, mais qu'au-delà, la *complexification* n'est pas une notion contradictoire. Or ce seuil de complexité, supposait-il, est aussi celui où la structure de l'objet devient plus simple que la description de ses propriétés. Dans le cas habituel, celui des machines simples, il est moins compliqué de dire avec des mots ce dont l'automate est capable que de présenter son plan de câblage. Pour des automates complexes, l'inverse serait vrai : il serait plus simple, voire infiniment plus simple, de décrire la structure de l'automate que de décrire complètement son comportement. C'est sur cet argument que von Neumann fit reposer sa critique du modèle de McCulloch et Pitts au symposium Hixon.

Elisabeth Pacherie nous a demandé, entre autres choses, de dire notre opinion sur le lien de solidarité entre fonctionnalisme et machine de Turing. Je crois que la question devrait être autre. Ce n'est pas de machine qu'il s'agit (Turing vs connexionnisme), mais de regard porté sur la machine. Von Neumann n'avait-il pas devant les yeux le théorème d'équivalence proposé par McCulloch entre la machine de Turing universelle et la première machine connexionniste ? Or ce qui retint son attention, et ce qui (selon Gödel lui-même) fonde sa conjecture, c'est la question de l'indécidabilité de l'arrêt: il n'y a pas de moyen plus économe de décrire le comportement de la machine que d'énumérer toutes ses productions possibles (une infinité). De tous les cybernéticiens, von Neumann aura été le seul à être sensible à cet aspect du travail de Turing, et de la logique formelle en général. Alors que les autres invoquaient à tout bout de champ la thèse de Turing (trop vite baptisée "théorème"), comme pour affermir leur foi dans la réductibilité du comportement humain au fonctionnement d'une machine, lui prenait la perspective inverse, et retenait la leçon proprement paradoxale que le comportement d'une machine n'est pas la chose banale que l'on croit puisque aucune machine ne peut le prévoir complètement.

La conjecture de von Neumann sur la complexité est peut-être le seul outil théorique qui puisse fonder un monisme matérialiste non réductionniste. Elle rend, par exemple, non contradictoires les deux propositions suivantes : 1) Des mécanismes physico-chimiques sont capables de produire la vie ; 2) La vie est (infiniment) plus complexe que les mécanismes physico-chimiques qui l'ont engendrée. Il est cohérent d'embrasser une ontologie non substantialiste (ici, non vitaliste) *et* non réductionniste.

Ce n'est paradoxalement pas dans les sciences cognitives que la conception neumanienne de la complexité a eu un impact important, c'est dans la philosophie sociale. J'évoquerai simplement le cas de Hayek, dont la position théorique peut être résumée par la conjonction des deux propositions suivantes, apparemment incompatibles, si ce n'était la conjecture de von Neumann: 1) Ce sont les hommes qui "agissent" leur société ; 2) La société leur échappe, parce qu'elle est (infiniment) plus complexe qu'eux. On peut sans incohérence être non réductionniste sans avoir à accepter le holisme<sup>1</sup>.

Hayek nomme "autotranscendance" du social par rapport aux individus cette configuration. Il est tentant de faire le rapprochement avec la figure phénoménologique de la "transcendance dans l'immanence", qui elle aussi vise quelque chose de l'ordre de la mise en extériorité de soi par rapport à soi. J'en viens à ma conjecture. J'ai suggéré ailleurs que la conception brentanienne, puis husserlienne de l'intentionnalité était beaucoup plus à même d'accueillir le thème de la complexité que la version qu'en ont donnée, non sans commettre un contresens radical, Chisholm, puis Quine<sup>2</sup>. Ce qui est en cause, dans l'absence de postérité de la conjecture de von Neumann en sciences cognitives, c'est donc le mariage entre ces dernières et la philosophie analytique de l'esprit — mariage qui est plus un accident de l'histoire qu'une nécessité.

Jean-Pierre DUPUY  
CNRS - École Polytechnique - CREA  
1, rue Descartes, 75005 Paris

\*\*\*

Olivier HOUDÉ. — *Psychologie cognitive expérimentale et controverse Fodor-Jacob à propos de « l'esprit atomique »*

Le point de vue philosophique original défendu par Pierre Jacob, avec conviction et talent, est que, contrairement à ce qu'affirme Fodor, le réalisme intentionnel n'est pas condamné à l'atomisme sémantique, ou conception d'un « esprit atomique ».

La raison de cette condamnation (où seules comptent les relations « pensées-monde » et non les relations entre pensées) serait, selon Fodor, que le holisme sémantique est incompatible avec la possibilité de lois psychologiques intentionnelles et dès lors avec la « tâche » même de la

<sup>1</sup> Cf. J.-P. Dupuy, *Introduction aux sciences sociales*, Ellipses, 1992.

<sup>2</sup> Cf. *Aux origines des sciences cognitives*, La Découverte, 1994.

psychologie scientifique ( et, plus généralement, des sciences cognitives ( qui consiste à établir ces lois. Le risque épistémologique est donc grand si l'on en croit Fodor, *mais fait-il réellement frémir le psychologue ?* A partir d'une série de remarques-questions relatives à l'article de Pierre Jacob, je m'attacherai à montrer que non. Ce sera une autre façon, plus directe et moins structurée, de répondre négativement à la question « un réaliste intentionnel est-il condamné à l'atomisme sémantique ? ».

(1) Le cœur de l'argumentation de Fodor, telle qu'elle est restituée par Jacob, est que si la propriété sémantique de la croyance d'un individu dépend de ses relations causales et/ou inférentielles avec les autres attitudes propositionnelles de l'individu (holisme sémantique), et si deux individus ne partagent pas *toutes* leurs attitudes propositionnelles, alors le risque est qu'aucune loi psychologique intentionnelle ne soit jamais instanciée par une paire d'individus. Le psychologue «cognitiviste» qui, comme c'est mon cas, étudie expérimentalement le développement et le fonctionnement des opérations logico-mathématiques (les capacités numériques, la catégorisation, le raisonnement, etc.), situées au niveau des « systèmes centraux » de Fodor, est frappé par la proximité entre cette argumentation et l'antipsychologisme de Frege et de Husserl. Ces derniers affirmaient, il y a fort longtemps, qu'en raison de leur objectivité et de leur caractère public, les lois logiques ne peuvent dériver de représentations subjectives et privées, et qu'en raison de leur nécessité, elles ne peuvent dériver de représentations variables d'un individu à l'autre. Or, il est reconnu, tant en psychologie cognitive (Houdé, 1993) qu'en philosophie de l'esprit (Engel, 1989), que ce type d'argumentation n'est plus de mise aujourd'hui. *Fodor est-il un « antipsychologue attardé » ?* La question est brutale, mais il semble bien que ce soit celle qui se pose si l'on en juge aux efforts déployés par Jacob pour montrer que le holisme sémantique (l'étude des relations entre symboles ou pensées, dont les opérations logico-mathématiques sont l'expression parfaite) est compatible avec la recherche de lois psychologiques intentionnelles.

(2) Conjointement à ce sentiment d'anachronisme épistémologique, la question qui traverse constamment l'esprit, à la lecture du texte de Jacob, est de savoir à quel niveau de processus cognitifs se situe exactement le débat ? S'il s'agit des processus dits «intelligents», «supérieurs» ou «centraux» (tels les opérations logico-mathématiques évoquées dans le point précédent : capacités numériques, catégorisation, raisonnement, etc.), la position de Fodor, dans «The Modularity of Mind», est claire : c'est à ce niveau qu'il situe la fixation des croyances et celle-ci est de type «isotrope et quinienne», c'est-à-dire compatible avec un holisme cognitif (position proche, à quelques nuances près, de celle défendue par Jacob). Or on connaît la «première loi de Fodor sur l'inexistence des sciences

cognitives », selon laquelle plus un processus est global (c'est-à-dire isotrope), moins on sait de choses à son sujet. Rien d'étonnant, dès lors, qu'il condamne le réalisme intentionnel, et son exigence de rigueur, à l'étude d'un « esprit atomique ». Le raisonnement de Fodor est cohérent, mais il ne peut convaincre le psychologue qu'à la condition, nécessaire et non suffisante, d'accepter une « focalisation scientifique » exclusive sur les systèmes périphériques de perception (notamment les phénomènes de vision) et de psycholinguistique. *Or qui aujourd'hui, en psychologie cognitive expérimentale, accepte une telle réduction ?* Le propos n'est pas de dire qu'il n'existe aucun psychologue répondant à ce profil, mais que leur nombre est limité et que la controverse « Fodor-Jacob » *n'est réellement palpitante que pour eux*. Pour les autres, défendre l'idée d'un « esprit anatomique », au sens de la considération des relations entre pensées, est un combat sans risque.

(3) Je voudrais illustrer cette question des niveaux de processus cognitifs en référence à la « tension », discutée par Jacob, entre l'atomisme sémantique et la propriété de « systématisme de l'esprit » définie par Fodor : un esprit est systématique au cas où s'il peut penser une proposition complexe, il aura la capacité de penser des propositions syntaxiquement reliées à la première. Jacob souligne que la systématisme, tout en étant formellement compatible avec l'atomisme sémantique de Fodor (en raison du caractère *syntactique* du rapprochement structural qu'elle établit), se combine plus harmonieusement avec l'atomisme sémantique. A la lecture de l'argumentation de Jacob sur cette « double compatibilité pondérée », le psychologue ne peut s'empêcher de rapprocher la « systématisme-combinée-avec-l'atomisme-sémantique » de ce que l'on appelle le « raisonnement par analogie ». Or ce rapprochement possible (qui me semble pertinent) introduit une rupture radicale, qui n'apparaît pas dans la discussion de Jacob, entre processus cognitifs de niveaux différents (en l'occurrence « périphériques » ou « centraux »). En effet, pour Fodor dans « The Modularity of Mind », le raisonnement par analogie est la plus pure expression de l'isotropie (holisme cognitif) et, en vertu de sa « première loi sur l'inexistence des sciences cognitives » citée plus haut, il affirme que « sur les processus très globaux comme le raisonnement par analogie, on ne sait rien du tout ».

Dans ce cas, à nouveau, soit conformément à l'affirmation de Fodor (qui ne manque pas de surprendre), la « systématisme-combinée-avec-l'atomisme-sémantique », suggérée par Jacob et entendue ici comme un processus de raisonnement par analogie, est susceptible d'être exclue du champ de la « focalisation scientifique », risque qui ne convaincra que peu de psychologues ; soit elle est mise en avant comme la propriété qui justifie l'étude scientifique d'un « esprit anatomique », ce qui est un

combat sans risque au regard de l'abondante littérature expérimentale sur le raisonnement par analogie.

(4) Le texte de Jacob m'inspire, par ailleurs, une remarque sur le projet de « naturalisation de l'intentionnalité ». Ce projet est présenté comme le premier objectif d'un réalisme intentionnel associé au physicalisme, position à laquelle souscrivent Fodor et Jacob, et qui consiste à s'efforcer de montrer comment les propriétés sémantiques des états mentaux d'un individu peuvent découler de propriétés non sémantiques du cerveau de l'individu. De ce point de vue et dans le cadre d'une réflexion à visée interdisciplinaire (il en va de la « tâche » de la psychologie et des sciences cognitives), je m'étonne de l'absence de référence, concernant la discussion du holisme cognitif ou sémantique, aux travaux de neuroscience cognitive sur les fonctions exécutives et intégratrices du cortex préfrontal, « organe de la pensée » selon l'expression de Changeux. Les études anatomiques, utilisant les techniques de marquage neuronal, indiquent que le cortex préfrontal est une « plate-forme d'intégration » qui (a) présente des connexions réciproques avec toutes les aires impliquées dans le traitement des informations sensorielles, (b) reçoit des afférences des structures limbiques (monde intéroceptif : état affectif et motivationnel), (c) est en relation étroite avec les structures de mémorisation d'expériences passées (complexe amygdalo-hippocampique, thalamus), et (d) participe au contrôle moteur par ses efférences sur le striatum et le cortex prémoteur. Autant d'éléments qui sont susceptibles d'en faire un véritable « carrefour anatomico-cognitif ». C'est ce que confirment les données de la neuropsychologie humaine où il apparaît que le cortex préfrontal est impliqué dans les fonctions exécutives centrales (Dubois *et al.*, 1994). Son rôle est celui d'un « superviseur attentionnel » qui intervient lorsque les systèmes « routiniers » de traitement dépendant de la force d'activation des processus perceptifs (tels les systèmes périphériques de Fodor) ne sont pas suffisants pour sélectionner les stratégies appropriées (Shallice, 1995). Selon Diamond (1991), le cortex préfrontal assure la mise en relation des informations dans l'espace et dans le temps *et inhibe* les tendances dominantes à l'action (voir pour une analyse des rapports entre cette fonction d'inhibition et le développement cognitif : Dempster & Brainerd, 1995 ; Houdé, 1995).

Quoi qu'en fasse le philosophe de l'esprit, et même si ces éléments de neuroscience cognitive ne peuvent être intégrés tels quels à son argumentation, il me semble que le réalisme intentionnel exige qu'ils soient « appelés à la barre » du procès de « l'esprit atomique ».

(5) La dernière remarque, enfin, est relative à la conclusion de l'article de Jacob, qui souligne la compatibilité entre la conception philosophique

d'un « esprit anatomique » et la conception psychologique actuelle d'une pensée dite « domaine-spécifique ». Ce rapprochement est intéressant, mais, contrairement aux apparences, on peut déplorer qu'il reste très fodorien dans son esprit, au sens où une stratégie de recherche réaliste consiste, selon Fodor, à n'étudier que les systèmes cognitifs modulaires et à accueillir la mort dans l'âme les arguments convaincants en faveur du caractère non modulaire d'un système. Quels que soient les états d'âme de Fodor, je pense que de tels arguments existent concernant la « pensée » (voir la remarque précédente sur les fonctions exécutives et intégratrices du cortex préfrontal) et qu'ils sont compatibles avec le projet d'une naturalisation de l'intentionnalité.

Les remarques consignées dans ce « texte-débat » témoignent du vif intérêt pris à la lecture de l'article de Pierre Jacob. Elles témoignent sans doute aussi (ici ou là) de la naïveté du psychologue face à des subtilités philosophiques qui lui échappent.

Olivier HOUDE

Université Paris V, LaPsyDEE - CNRS  
GDR « Sciences cognitives de Paris »  
46 rue Saint-Jacques, 75005 Paris

#### **BIBLIOGRAPHIE**

- Dempster, F.N, & Brainerd, C.J. (1995). *Interference and inhibition in cognition*. New York: Academic Press.
- Diamond, A. (1991). Neuropsychological insights into the meaning of object concept development. In S. Carey and R. Gelman (Eds.). *The epigenesis of mind: essays on biology and cognition* (pp.67-110). Hillsdale, N.J.: Lawrence Erlbaum.
- Dubois, B., Pillon, B., & Sirigu, A. (1994). Fonctions intégratrices et cortex préfrontal chez l'homme. In X. Seron et M. Jeannerod (Eds.). *Neuropsychologie humaine* (pp. 453-469). Bruxelles: Mardaga.
- Engel, P. (1989). *La norme du vrai : philosophie de la logique*. Paris: Gallimard.
- Houdé, O. (1993). La référence logico-mathématique en psychologie : entre méthode universelle et rationalité arrogante. In O. Houdé & D. Miéville (Eds.). *Pensée logico- mathématique : nouveaux objets interdisciplinaires* (pp. 47-119). Paris: PUF.
- Houdé, O. (1995). *Rationalité, développement et inhibition*. Paris: PUF.
- Shallice, T. (1995). *Symptômes et modèles en neuropsychologie*. Paris: PUF.

\*\*\*

Daniel MEMMI. — *Le connexionnisme est-il réductionniste ?*



Les modèles connexionnistes sont actuellement une des principales alternatives à l'approche symbolique classique (illustrée notamment par l'Intelligence Artificielle, la psychologie cognitive et la linguistique formelle). Comme le connexionnisme s'inspire clairement de processus neurobiologiques, on pourrait penser qu'il est réductionniste. En cherchant à modéliser la cognition par des réseaux neuronaux sans s'arrêter au niveau symbolique, le connexionnisme semble fondamentalement opposé au fonctionnalisme classique.

La réalité est autrement plus complexe. Tout d'abord, il y a maintenant une grande variété de modèles connexionnistes, et il faudrait à mon avis distinguer au moins trois grandes approches dans ce domaine. Une bonne partie des réseaux neuronaux sont développés en vue d'applications techniques (notamment en reconnaissance des formes), sans véritable prétention biologique ou psychologique (c'est aussi le cas des modèles neuronaux issus de la physique mathématique). Une autre approche s'efforce à une modélisation biologique fine et réaliste ; malheureusement les phénomènes traités sont encore généralement de trop bas niveau (comme souvent en neurobiologie) pour avoir une signification cognitive. Enfin une troisième voie consiste en la simulation de processus cognitifs proprement dit (comme la perception, le raisonnement, le langage...). C'est cette dernière direction seulement (le "connexionnisme cognitif") qui est vraiment pertinente dans ce débat, même s'il y a bien sûr de nombreuses interactions entre les différentes approches.

Or que peut-on constater au sujet de ce connexionnisme cognitif ? Il ne renonce pas en fait aux notions classiques : représentations, règles, catégories et structures. Ces entités abstraites informent la modélisation des tâches, le choix des architectures, le codage des entrées et des sorties, l'interprétation des résultats. Puis les chercheurs connexionnistes ont utilisé avec succès des techniques d'analyse de données (analyse hiérarchique, analyse factorielle...) pour extraire les classes et structures qui sont implicites dans le fonctionnement d'un réseau neuronal après apprentissage. On observe alors sans grande surprise que ces représentations implicites correspondent aux régularités de la tâche modélisée, telles que les psychologues et linguistes les avaient déjà dégagées et formulées...

Pour donner un exemple précis, un réseau récurrent multicouches peut apprendre (par rétropropagation) à reconnaître ou bien à interpréter des phrases à partir d'un corpus d'exemples. On montre ensuite, par analyse de données sur l'ensemble des états internes du réseau, que celui-ci a développé une grammaire implicite, avec des classes syntaxiques hiérarchisées correspondant au corpus d'apprentissage. Cette grammaire

systematique et structurée n'est visible que dans l'espace abstrait des états du système, mais elle est parfaitement compréhensible par un linguiste. On retrouve donc des concepts classiques dans ces modèles (ce qui rend les critiques de Fodor & Pylyshyn contre le connexionnisme en grande partie caduques).

Il y a cependant deux différences essentielles avec les modèles symboliques. Premièrement les entités classiques, dans la mesure où on les retrouve dans les modèles connexionnistes, ne sont pas directement causales dans ces modèles. Ces entités permettent de décrire et d'expliquer le comportement des réseaux au niveau symbolique, mais la causalité effective du système se situe à un niveau plus bas, celui des connexions et neurones formels. Ce niveau connexionniste est d'ailleurs lui-même fortement idéalisé et simplifié par rapport à la neurobiologie réelle. Deuxièmement, du fait de cette causalité plus fine, les concepts classiques ne fournissent que des approximations du fonctionnement des réseaux. Il ne s'agit pas d'une simple compilation connexionniste de modèles symboliques. Les systèmes neuronaux offrent ainsi des qualités nouvelles par rapport aux modèles classiques : résistance au bruit, traitement flou, résolution de contraintes souples, apprentissage graduel...

Autrement dit, ni le fonctionnalisme ni le réductionnisme ne paraissent vraiment appropriés (du moins dans leur versions extrêmes) pour décrire cette pratique scientifique. Le connexionnisme cognitif montre comment les concepts du niveau symbolique restent utiles et même inévitables lorsqu'ils correspondent aux régularités du domaine. En ce sens le fonctionnalisme semble indépassable : le niveau psychologique de description reste le plus clair et le plus économique pour la modélisation cognitive. Mais les modèles connexionnistes montrent en même temps comment les notions classiques sont ré-interprétées et remaniées au sein d'une nouvelle théorie plus fine, qui explique de nouveaux phénomènes. On voit donc aussi l'intérêt d'une approche réductionniste, comme stratégie de recherche plutôt que comme dogme, d'autant plus que les modèles neuronaux n'arrivent encore généralement pas vraiment au niveau biologique proprement dit.

Et le connexionnisme ne constitue qu'un exemple parmi d'autres de la complexité des rapports entre théories concurrentes ou successives dans un domaine scientifique. Même dans les cas où une réduction de principe a pu être obtenue, la théorie de plus haut niveau est souvent non pas remplacée totalement, mais ré-interprétée et modifiée par la nouvelle théorie, qui est elle-même profondément informée par la première. On constate fréquemment dans les sciences modernes des interactions entre théories plutôt qu'une guerre totale. Ainsi la thermodynamique est en

principe réductible à la mécanique statistique (la température par exemple se réduisant à l'agitation des molécules) mais on continue à utiliser les concepts de la thermodynamique, même si nous les comprenons maintenant différemment. De même la génétique et la biologie moléculaire continuent à s'influencer mutuellement, et le concept abstrait de gène a été grandement enrichi, mais non remplacé par la découverte de l'ADN et les progrès remarquables de la biochimie.

On pourrait alors se demander ce qui motive la rigidité de certaines positions théoriques dans les sciences cognitives (en dehors des luttes d'influence habituelles !). Il s'est probablement produit une certaine réification du niveau d'explication et de causalité choisi dans chaque théorie. Ce niveau devient tellement naturel et évident pour les praticiens d'une approche donnée qu'il devient très difficile de le remettre en question. Mais c'est oublier que nous n'avons affaire qu'à des modèles provisoires, qui sont de plus particulièrement imprécis et sous-déterminés dans nos domaines. Il ne me paraît donc pas raisonnable de privilégier exclusivement une causalité symbolique, ou connexionniste, ou biologique, ou physique... En bref, les niveaux d'explication ne sont pas uniques, mais complémentaires, et chacun peut apporter sa contribution à l'étude de la cognition.

Daniel MEMMI  
LIFIA-IMAG  
46 avenue Félix Viallet  
38000 Grenoble  
memmi@imag.fr

**BIBLIOGRAPHIE**

- Bechtel W. & Abrahamsen A. (1991) *Connectionism and the Mind*, Basil Blackwell.
- Burnod Y. (1988) *An Adaptive Neural Network: The Cerebral Cortex*, Masson.
- Churchland P.S. (1986) *Neurophilosophy*, MIT Press.
- Clark A. (1989) *Microcognition*, MIT Press.
- Crucianu M. (1994) *Représentations Structurées dans les Réseaux Connexionnistes*, Thèse Université de Paris-Sud, Orsay.
- Elman J.L. (1990) Finding structure in time, *Cognitive Science* 14 (2).
- Elman J.L. (1991) Distributed representations, simple recurrent networks, and grammatical structure, *Machine Learning* 7.
- Fodor J.A. & Pylyshyn Z.W. (1988) Connectionism and cognitive architecture : a critical analysis, *Cognition* 28 (1-2).
- Hinton G.E. ed. (1990) Connectionist Symbol Processing, special issue, *Artificial Intelligence* 46 (1-2).
- Memmi D. & Visetti Y.M. eds. (1990) Modèles Connexionnistes, n° spécial, *Intellectica* 9-10.
- Memmi D. (1993) Representations in connectionist networks, *Proceedings of Neuro-Nîmes 93*, Nîmes.
- Rumelhart D.E. & McClelland J.L. eds. (1986) *Parallel Distributed Processing*, MIT Press.
- Smolensky P. (1988) On the proper treatment of connectionism, *Behavioral and Brain Sciences* 11 (1).

\*\*\*

Jacques PITRAT. — *L'intelligence dépend des caractéristiques de son support matériel*

Pour le fonctionnalisme, seules les fonctions remplies par un système traitant l'information ont de l'importance ; sa nature physique n'est pas essentielle. Ceci est peut-être exact sur le plan théorique, mais en pratique une tâche doit être accomplie dans un temps raisonnable. Aussi les capacités d'un système seront limitées si celui-ci est mis en œuvre sur un support privé de certaines possibilités. Nous allons prendre pour exemple l'intelligence humaine, irrémédiablement limitée parce que son support est notre cerveau. Ces limitations ont deux origines : la structure de notre cerveau et sa composition à base de neurones. Par contre, certains systèmes artificiels ne sont pas soumis à ces limitations.

*Limitations dues à la structure de notre cerveau.* — Il serait possible de concevoir des cerveaux plus puissants que les nôtres simplement en y incluant plus de neurones ou en les câblant autrement.

En particulier, le nombre de nos éléments actifs nous limite : nous en avons beaucoup, de l'ordre de dix milliards, mais nous ne pouvons en avoir plus. Nous avons des zones spécialisées dans la perception visuelle ou le langage ; avec davantage de neurones, nous pourrions avoir en plus des zones spécialisées dans les mathématiques ou l'informatique. L'évolution n'a pas encore jugé bon de nous réserver des zones spécialement adaptées à ces activités. Les systèmes artificiels sont loin d'avoir pour le moment autant d'éléments actifs que nous, mais ce nombre progresse très vite ; on ne voit pas pourquoi ils seraient contraints de s'arrêter à notre niveau.

Nous avons au plus une dizaine de mémoires de travail qui nous permettent de stocker des résultats intermédiaires pendant que nous réfléchissons. Cela nous gêne par exemple si nous essayons d'énoncer ou de comprendre des structures syntaxiques très imbriquées. Or trois imbrications nous posent déjà un problème quand nous entendons la phrase suivante : *Le bébé que le chien que le père que la femme bat mord taquine ronge un os*. Nous n'arrivons pas à stocker les sujets au fur et à mesure pour les retrouver quand nous rencontrons le verbe correspondant. Bien qu'elle soit très sévère, nous avons pourtant peu conscience de l'existence de cette limite parce que nous nous y sommes adaptés. Comme nous y sommes tous soumis, nous nous exprimons en en tenant compte : nous ne mettons jamais nos auditeurs dans une situation telle qu'elle les empêcherait de nous comprendre. Pour la phrase précédente, nous n'avons aucun mérite, car nous n'arrivons pas plus à l'engendrer qu'à la comprendre. Cette limitation n'intervient pas uniquement dans nos rapports avec le langage, mais chaque fois que nous faisons de tête un raisonnement complexe. Nous sommes obligés de prendre une mémoire auxiliaire comme une feuille de papier ; cela nous ralentit énormément. Il semble possible de concevoir un cerveau qui aurait une mémoire de travail de plus grande capacité, mais ce ne serait plus un cerveau humain. La plupart de nos programmes informatiques ne pourraient fonctionner avec une restriction aussi sévère.

Par ailleurs, la conscience réflexive nous permet d'avoir accès à une partie de ce qui se passe dans notre cerveau. Elle nous est très utile, par exemple pour comprendre les raisons de nos succès ou de nos échecs ; nous pouvons ainsi apprendre ce qui est bien et éviter de reproduire nos erreurs. Mais cette conscience est trop limitée : certaines activités sont totalement inconscientes, par exemple la perception visuelle. Il n'est ni bon ni possible de s'observer totalement, mais un système informatique peut observer ce qu'il veut parmi ce qu'il fait. Nous aurons un jour des systèmes artificiels qui nous seront supérieurs simplement parce qu'ils sauront mieux que nous les raisons de leurs actions.

*Limitations dues à l'utilisation de neurones.* — D'autres limitations viennent de ce que le fonctionnement de notre cerveau repose sur le neurone. Or le neurone manque de qualités présentes dans d'autres éléments actifs comme ceux utilisés actuellement dans les ordinateurs.

Ainsi le neurone fonctionne lentement : les temps s'expriment en millisecondes alors que, pour les ordinateurs actuels, nous utilisons la nanoseconde, soit un million de fois plus vite. Une vie humaine de 80 ans accélérée un million de fois se réduirait à 42 minutes. Et la vitesse des ordinateurs continue sa progression sans faiblir ! Pour le moment, nous n'utilisons pas pleinement cet avantage parce qu'ils ne fonctionnent pas suffisamment en parallèle. Mais nous ne sommes pas obligés à nous restreindre à des architectures de type von Neumann. Le jour où nous saurons réaliser des machines massivement parallèles, leur puissance de traitement dépassera largement celle de notre cerveau.

En outre, un avantage essentiel de la technologie actuelle des ordinateurs est la possibilité d'y charger de nouvelles informations et inversement de décharger celles qui y sont contenues. Cette activité est totalement interdite aux éléments de notre cerveau, nous ne pouvons ni extraire son contenu, ni le changer à volonté. Cet avantage est constamment utilisé par l'informaticien. Il permet en particulier de multiplier les systèmes qui donnent de bons résultats. Certains logiciels sont dupliqués à des millions d'exemplaires ; chaque nouvelle copie n'a qu'un coût minime comparé à celui qu'a demandé la réalisation de l'exemplaire de départ. Il est par ailleurs facile de les améliorer : s'ils cessent d'être les meilleurs, on les remplace instantanément par le nouveau meilleur système. Pour l'homme, l'oubli est souvent aussi difficile que l'apprentissage.

La possibilité précédente permet ainsi de sauvegarder l'état d'un système pour le restaurer en cas de destruction accidentelle. Elle assure donc aux systèmes informatiques une quasi-immortalité : comme ils peuvent être reproduits sans limite, il est peu probable que tous les exemplaires soient simultanément détruits. Cette augmentation de durée de vie assure le maintien des connaissances, alors que l'on peut comparer la mort d'un vieillard à la disparition d'un dictionnaire unique.

Je présente encore dans (Pitrat 1995) d'autres limitations de notre cerveau auxquelles les systèmes informatiques ne sont pas plus soumis qu'à celles que nous venons de voir. Malheureusement, nous ne savons pas encore bien tirer parti des avantages des ordinateurs parce qu'il n'existe pas de modèle qui puisse nous inspirer : aucune forme d'intelligence connue dans l'univers n'utilise ces possibilités. Notre plus gros handicap vient de

ce que des humains ont à mettre en œuvre des systèmes extrêmement complexes alors que leur intelligence est si limitée. C'est de là que vient le seul argument sérieux contre les espoirs de l'Intelligence Artificielle : théoriquement, il est possible de réaliser des systèmes qui feront preuve d'un comportement plus intelligent que le nôtre, mais nous n'y arriverons jamais parce que nous ne sommes pas assez intelligents pour les réaliser. Je pense que l'avenir donnera tort à cet argument. Mais, même si nous atteignons un jour notre but, c'est une tâche de très longue haleine ; aussi ne devons-nous pas nous attendre à des résultats spectaculaires dans un proche avenir.

Notre seule chance d'échapper aux limitations inhérentes au cerveau, support matériel de notre intelligence, est de développer des supports artificiels qui ne sont pas soumis aux mêmes restrictions.

Jacques PITRAT  
CNRS-LAFORIA  
IBP-Université Pierre et Marie Curie  
Paris

### **BIBLIOGRAPHIE**

Pitrat J. (1995) De la machine à l'intelligence, Hermès.

\*\*\*

François RASTIER. — *Onze questions sur le fonctionnalisme.*

Même après avoir lu la limpide introduction d'Elisabeth Pacherie, il n'est pas facile d'entrer dans le débat entre fonctionnalistes. En béotien, je dirai ici pour quelles raisons il me paraît faussé, et pourquoi il ne semble pas nécessaire d'y entrer. D'une part, ses termes me paraissent présupposer des thèses inacceptables, dans la mesure où elles restent indiscutées ; d'autre part, les moyens qu'il emploie vont souvent à l'encontre de ses objectifs affichés.

Le fonctionnalisme connaît mille nuances. Je discuterai ici essentiellement les thèses du fonctionnalisme de *mainstream*, celui qui a servi de théorie unifiante aux sciences cognitives.

*Etats mentaux.* — Qu'est-ce qu'un état mental, et qu'est-ce qui distingue un état A d'un état B ? Comment l'esprit connaîtrait-il des états discrets, clairement délimités, identiques à eux-mêmes dans leur durée ?

Quelle est leur métrique ? Rien de ce que l'on connaît du cerveau en activité ne permet de dire qu'il est à un moment  $t$  dans un et un seul état X ou Y. On m'objectera certes que je confond le niveau computationnel et le niveau de l'implémentation, mais c'est comme on verra une distinction que je récusé.

Sans base empirique, la notion d'état mental ne serait-elle pas tout simplement une réification parmi d'autres du positivisme logique, fondement idéologique de la philosophie de l'esprit ? On connaît l'atomisme de cette tradition, lié à son réalisme des individus : elle a déjà produit la notion étrange *d'état de choses*. Cet atomisme est indispensable ici pour soutenir le parallélisme entre l'esprit-cerveau et la machine de Turing : cet automate abstrait connaît en effet des états discrets se succédant strictement.

*Fonctions.*— Le concept de fonctionnalisme est ambigu, dans la mesure où il réfère tantôt aux fonctions des états mentaux, tantôt au fonctionnement de l'esprit-cerveau. On sait d'ailleurs que dans la tradition fonctionnaliste, le terme de fonction a dérivé peu à peu du sens algébrique à son acception biologique (dans le téléofonctionnalisme d'aujourd'hui).

A quelle condition un état mental peut-il être dit revêtir une fonction, ou un rôle fonctionnel, dit aussi rôle causal ? Dans la mesure où il en cause un autre. On reconnaît là une involution mentaliste du behaviourisme réductionniste (mais non éliminationniste).

Il suffira de montrer que la fonction s'exerce par un calcul (une computation) pour permettre des parallèles insistants entre esprit et Machine de Turing, cerveau et ordinateur (cf. Putnam, 1994, p. 507 : « le fonctionnalisme soutient que nous sommes semblables à des ordinateurs et que nos états psychologiques sont simplement des "états fonctionnels" »).

*Causes.* — Quelle est la valeur de la causalité dans l'explication des processus mentaux et cérébraux ? Outre qu'un même effet peut évidemment avoir plusieurs causes, et une même cause plusieurs effets, et que l'on n'a pas oublié le rôle de l'hypothèse causale dans le behaviourisme, l'usage fonctionnaliste du concept de cause permet de promouvoir une image mécaniste des systèmes physiques et biologiques. Ils restent décrits en termes de conditions initiales (qu'on appelle des causes), et non d'instabilité structurelle.

Cette image a sous-tendu les programmes positivistes à la Taine et soutient encore le programme du fonctionnalisme. Si elle vaut encore pour la macrophysique de sens commun (préscientifique) elle ne s'accorde cependant pas ni avec les conceptions contemporaines de la causalité : ni celles de la physique, ni celle des sciences humaines (dans leur effort pour formuler des théories de l'action).



Pourquoi au demeurant l'effet se connaîtrait-il tout entier dans sa cause ? En quoi la conception étiologique de la connaissance scientifique aurait-elle une valeur explicative ? Elle est traditionnelle dans l'aristotélisme ; elle a connu un nouvel essor avec la physique newtonienne, mais elle a été ruinée par la physique contemporaine. Dans les sciences sociales, nous n'avons pas accès à des causes, mais à des conditions — dont nul ne peut dire qu'elles soient initiales.

Le réductionnisme a pour principe d'une part de privilégier exclusivement un niveau de description, et d'autre part de confondre les corrélats avec des causes. Par exemple, la description biochimique d'un anticorps n'explique en rien sa formation et son efficacité. La balistique oculaire n'explique pas la lecture, elle en est un corrélat moteur. Ou encore, les corrélats neuronaux des "états mentaux" n'expliquent pas leur cours phénoménologique.

Rien n'échappe en droit à la description scientifique, à condition qu'elle tienne compte de la complexité de ses objets, et, passées les simplifications qu'exige la définition de niveaux d'analyse, elle permette de restituer cette complexité.

*Un étrange monisme.* — On connaît la thèse objectiviste du monisme cognitif. Comme par définition l'objet est indépendant de l'esprit, que cette indépendance fonde et permet la séparation sujet / objet, que la matière est le seul lieu de l'objectivité, le fonctionnalisme pose évidemment l'objectivité de l'esprit. Il doit donc traiter les états mentaux comme des objets, conformément aux principes d'une psychologie antisubjectiviste. Comment cependant relier l'objectivité de l'esprit (décrit par une psychologie formelle) et l'objectivité physique ?

On a souvent souligné le dualisme latent du fonctionnalisme (l'auteur, 1991 ; Putnam 1994 ; Dupuy, ici-même ; Bachimont, à paraître). Il procède sans doute de la contradiction originaire du positivisme logique, qui entend tout à la fois être empirique et formel.

On se souvient que Chomsky se plaçait lui-même dans la lignée de la linguistique cartésienne. Bien qu'aujourd'hui le dualisme cartésien ne soit sérieusement défendu par personne, c'est pourtant lui que le monisme fonctionnaliste se propose d'annuler en le reformulant, la matière et l'esprit devenant le physique et le symbolique.

La distinction entre ces deux instances fait problème, si l'on admet entre elles un rapport de représentation. Alors que chez Descartes l'Intellect Archétype était, comme le dit Cassirer, « le lien fixe, le crampon de fer qui retient unis la pensée et l'être, la vérité et la réalité », chez Fodor son absence conduit au "solipsisme méthodologique", et l'on sait bien que le solipsisme est une position typique de l'idéalisme. Cette aporie sémantique sera "résolue" par la voie syntaxique. Voici comment.

Pour annuler le dualisme matière / esprit, on réarticule une autre forme de dualisme, le dualisme aristotélicien de l'opposition forme / substance. Certes, on emploie le mot *substrat*, mais il n'est pas défini autrement que la substance — concept qui ne joue plus, comme l'a souligné Weyl, aucun rôle en physique. En fait, on appelle *substrat* ce que l'on décide de considérer comme inessentiel.

Le dualisme forme / substance est une sorte de la différence ontologique entre l'essence et l'existence, philosophiquement beaucoup plus importante que l'opposition entre la matière et l'esprit. On masque ordinairement ce dualisme en considérant que la forme est immanente à la substance, tout en répliquant la division dualiste au sein de la matière elle-même.

*Types et occurrences.* — On postule d'abord — sans nulle base empirique — que la pensée est faite d'opérations sur des symboles (au sens logique du terme). Le rapport de la forme et de la substance se joue alors tout entier dans la mise en scène des rapports entre symboles-types et symboles-occurrences. Le type est idéal, l'occurrence est matérielle : comme le programme mental ne tient compte que de la forme des symboles, et non bien entendu du substrat matériel où sont inscrites les occurrences, il va de soi que toutes les occurrences du même type sont exécutées de la même manière (cf. Bachimont, à paraître). Or les symboles dans leur forme, et en tant que types, sont l'objet de la computation, alors que dans leur inscription matérielle, et en tant qu'occurrences, ils font l'objet de l'implémentation. Il suffit alors de dire que le rapport matière / esprit est homologue du rapport implémentation / computation, pour avoir apparemment résolu le problème de leur articulation. Cette solution subtile — au sens où elle subtilise le problème en paraissant le résoudre — repose d'une part sur l'identité computationnelle des types et des occurrences, d'autre part sur la thèse (traditionnellement idéaliste) de l'indépendance des formes à l'égard de leurs substrats.

Le rapport entre les types et les occurrences est un problème crucial en sémiotique et en sémantique : il régit par exemple les questions de la polysémie et de la typicalité, et plus généralement de la catégorisation. Or, caractériser ce rapport, *qualifier comment les occurrences diffèrent des types, c'est là un objectif majeur de l'interprétation*. Postuler qu'elles ne diffèrent pas, c'est estimer qu'elle va de soi, façon nonchalante de la nier. Ainsi, comme toute idéologie formaliste, le fonctionnalisme repose sur le suspens de l'interprétation — que permet au demeurant sa sémiotique limitée aux seuls symboles logiques.

*Le physicalisme de sens commun.* — A la différence du matérialisme vulgaire, le physicalisme ne suppose pas par lui-même la croyance en

l'existence d'objets discrets, nombrables, et pourvus chacun d'une identité à soi. Parmi les positions théoriques ouvertes par le positivisme, pourquoi le fonctionnalisme a-t-il choisi de poursuivre dans la voie d'un physicalisme de sens commun ? Ce choix ne procède pas d'une croyance aveugle, et Quine par exemple soutiendrait que les objets macroscopiques n'ont rien de fondamental, mais que le privilège qu'on leur accorde est essentiellement pratique. Cette résignation plus ou moins allègre au physicalisme de sens commun nous paraît avoir divers enjeux qui la rendent en fait inévitable :

(i) Ce physicalisme est simplifiant, et la réduction de la complexité physique est nécessaire au fonctionnalisme cognitif. Notamment, la métaphore computationnelle impose que la vie mentale soit décrite comme une succession d'états mentaux atomiques. Il peut sembler aujourd'hui abusif d'associer fonctionnalisme et atomisme logique : mais la filiation reste indiscutable.

(ii) La philosophie du langage (de Frege à Kripke) se caractérisait par son réalisme militant, c'est pourquoi elle a pris le nom, et notamment le nom propre, pour parangon de toute désignation. La philosophie de l'esprit en a hérité son réalisme des individus. Or, le physicalisme de sens commun est d'autant mieux compatible avec ce nominalisme que la notion d'objet individuel est précisément un produit typique de la physique de sens commun.

(iii) Le positivisme logique, par son programme même, suppose de pouvoir discerner des objets empiriques entre lesquels découvrir ou établir des relations logiques. Son empirisme est donc, dans la tradition de l'empirisme lockien, une forme d'atomisme. L'analyse logique appelée par le Manifeste de 1929 ne peut on le sait traiter ni du continu, ni de l'infini, ni des très grands nombres : elle ne peut donc s'exercer que sur des objets de sens commun.

Tout cela a rendu de fait indissoluble le lien entre l'objectivisme cognitif et la croyance propre au physicalisme de sens commun que le monde est composé d'objets macroscopiques.

*Les guises de la matière.* — Le fonctionnalisme est un formalisme qui voudrait penser son inscription matérielle. Il affronte à sa manière la contradiction propre au positivisme logique entre un empirisme (matérialiste) et un formalisme (idéaliste). Il ajoute à cela l'objectif de trouver un fondement positif à la logique, fonder la Raison en nature. Or :

(i) Si l'esprit est un système matériel, de quelle sorte de matérialité s'agit-il ? La question ne sera pas posée, car le seul régime de matérialité admis par le positivisme logique est celui que décrit la physique. Ce privilège sous-tend au demeurant la thèse de l'unité de la science.

(ii) Dans l'argumentation fonctionnaliste, le concept philosophique de matière est confondu avec le concept physique de matière, ce qui donne à ses conclusions métaphysiques l'apparence de conclusions scientifiques.

(iii) La thèse de l'indépendance des formes à l'égard des substrats est un des fondements de notre tradition métaphysique (cf. Aubenque, 1962). C'est précisément elle qui est à la racine de la thèse de l'indépendance de l'âme et du corps. On peut s'étonner que la philosophie de l'esprit, occupée à combattre le spiritualisme, ne soit pas en mesure, sur ce point, de discerner son propre idéalisme.

Cette thèse d'indépendance des formes et des substrats reste à la base du programme de l'IA forte. Nous voici devant un matérialisme qui ne se soucie pas des substrats matériels. Outre d'un mécanisme, il s'agit d'un mécanisme : seul un artefact peut changer de substrat matériel (les premiers couteaux étaient en pierre, les premières machines à calculer en bois).

(iv) Cette matière abstraite est un concept philosophique. Si on ne veut pas lui opposer le concept physique, on doit cependant remarquer que ce concept philosophique ressemble fort à celui d'une physique périmée.

*L'orientation métaphorique.* — Le programme de naturalisation conduit à l'explication du vivant en termes de causalité physique. Il renvoie aux sciences physiques, et non à celles de la vie. La naturalisation est en fait une physicalisation, conformément au programme du positivisme logique.

Il s'agit certes d'un processus plus général, et il faudrait s'interroger à ce propos sur le rôle idéologique de la biochimie. D'autant plus que le passage métaphorique de l'animé à l'inanimé reste caractéristique du mécanisme ancien et moderne (de La Mettrie à Changeux).

Cette orientation métaphorique, de l'ordre du désir, conduit de fait à négliger les propriétés caractéristiques de la matière organique, comme par exemple la dissymétrie moléculaire mise en évidence par Pasteur. Certes, comme le disait fortement Lénine, la matière pense. Mais il est oiseux d'étendre ce propos à la matière inorganique. L'insistance réitérée sur "les systèmes naturels et artificiels", les "machines qui pensent" et "l'esprit qui compute" réalise pourtant cette extension, avec les puissants moyens de la rhétorique binaire.

Il serait déjà peu intéressant comme le remarque Auroux, de comparer un marteau à un poing, sous prétexte qu'ils peuvent casser des noix. Il serait moins palpitant encore de décrire le poing comme une sorte de marteau. Ou encore, comme le demande Jaynes, si l'on peut comparer à Venise une carte de Venise, que signifierait la comparaison inverse ? La carte peut être fidèle, le territoire ne l'est pas.

On oublie enfin que les systèmes artificiels, en qualité d'artefacts, appartiennent à la sphère du sémiotique, et ne sont pas des objets physiques comme les autres.

*Pour ne pas conclure.* — Le fonctionnalisme a été une source d'hypothèses, mais reste inaccessible à l'épreuve empirique. Sans se donner vraiment la peine de le réfuter, Putnam dit avoir abandonné le fonctionnalisme qu'il avait théorisé le jour où il s'est avisé de ceci : « même si Chomsky a raison, nous ne pourrons jamais le découvrir » (1994, p. 11).

Parce qu'il est irréfutable, le fonctionnalisme ne peut être qu'abandonné. Ses hypothèses constituantes se révéleront sans doute des préjugés, invérifiables parce que métaphysiques. Il garde cependant un avenir, car il est de l'ordre de la croyance, et le scientisme qui l'inspire reste une superstition si ordinaire aujourd'hui qu'elle paraît insouçnable.

*Pour des sciences cognitives sans ontologie.* — Le fonctionnalisme, issu du positivisme logique, reste tributaire de son ontologie : la matière est ce qu'il y a à connaître et constitue l'objet de toute science. Conformément au réalisme des individus de tradition nominaliste, la matière se distribue en objets discrets et pourvus d'une identité à soi. La notion d'états de choses concrétise cette préconception donnée pour évidente parce que posée *a priori*, mais pour moi opaque.

La doxa positiviste voudrait qu'on puisse découpler l'ontologie (description de ce qui est) de la métaphysique (qui traiterait de catégorisation des objets). Mais l'ontologie objectiviste préjuge précisément de l'existence et de la légitimité absolue de la catégorie générale d'objet. Dire que la matière se distribue en objets et en états de choses est déjà une position métaphysique, tout à fait suprême par son universalité. Elle reflète certes une macrophysique de sens commun, du millimètre au kilomètre, celle d'un aristotélisme affadi et oublié. Mais on ne saurait oublier que l'objet, considéré comme donné à une *simplex apprehensio* que le Cercle de Vienne appelait l'*évidence*, reste un *préjugé macroscopique* (selon la forte expression de Gonsseth).

De ce point de vue le programme de naturalisation du sens, que Sperber nomme à bon droit le *Graal* des sciences cognitives, ne parvient qu'à naturaliser la métaphysique. La matière sert de caution muette à ce programme — bien que les physiciens, de Heisenberg à Cohen-Tannoudji, soient rarement physicalistes.

En somme, il y a autant d'ontologies qu'il y a de sciences : chacune, par diverses procédures ou rituels d'objectivation, crée et recrée son objet, établissant peu à peu son domaine d'objectivité. En outre, au cours de leur pratique, les scientifiques se montrent heureusement opportunistes en ce qui concerne l'ontologie, et leur absence de prudence fait plaisir à voir.

Bref, faire de la science, ce n'est pas vérifier une ontologie, c'est se priver d'ontologie préconçue, et cette privation me paraît constituante, dans la mesure où elle permet de constituer les objets scientifiques en passant outre à l'objectivisme de sens commun.

*La question sémiotique.* — Les sciences sémiotiques, notamment, dans la mesure où elle déploient une activité critique, ont un effet dissolvant sur l'ontologie. D'une part le régime de matérialité des signes est spécifique, car les stimuli sémiotiques ne sont pas perçus ni traités comme les autres (ce qu'illustre par exemple le phénomène de la perception catégorielle) : les signes sont irréductibles à des événements physiques "comme les autres" (malgré Morris, qui définissait le signe comme un événement physique), et ce texte n'est pas fait que de traces noires<sup>1</sup>.

Ensuite, l'analyse des structures sémantiques montre que les différents genres de textes induisent des impressions référentielles diverses. Ces effets de réel doivent être étudiés si l'on veut rendre compte du caractère culturel de la cognition humaine.

Corrélativement aux contraintes qu'ils exercent sur la constitution des impressions référentielles, les textes suscitent des effets de réel. Chaque type de mimésis est un mode de production de l'être : en d'autres termes, le réalisme compulsif de la tradition occidentale les pourvoit d'une mission ontogonique réifiée en fonction référentielle.

Les textes scientifiques n'échappent point à cette règle : chaque discipline s'est pourvue de techniques éprouvées pour objectiver ainsi le champ sémantique dont elle édifie la théorie. La littérature fonctionnaliste pourrait être étudiée de ce point de vue. À l'atomisme logique correspond à la fois un émiettement de textes cités en arguments, des courants de pensée en noms propres, des théories en articles, et des articles en apologues (expériences de pensée : la terre-jumelle, la chambre chinoise) ou anecdotes (les segments de lapin). La longueur et l'inélégance souvent affectée mettent en scène le sérieux de l'objectivité, la complexité de l'argumentation, la précision scientifique. L'histoire de la pensée devient le présent d'une école, sa diversité l'espace clos d'une discussion entre collègues : ainsi se crée un *état de choses* philosophique.

Ces remarques de style, rapides et sommaires, non-pertinentes voire impertinentes, ne sont certes que celles d'un linguiste incapable d'argumenter en philosophe ; mais la philosophie de l'esprit est aussi un style.

François RASTIER  
CNRS, Centre de Linguistique Française

---

<sup>1</sup> L'*encrage cognitif*, selon certains mauvais plaisants.

**BIBLIOGRAPHIE**

- Aubenque, P. (1962) *Le problème de l'Être chez Aristote*, Paris, PUF.
- Bachimont, B. (1996) *Artéfacture : des machines qui pensent aux machines qui donnent à penser — Critique du formalisme en intelligence artificielle*, Thèse, EHESS.
- Frege, G. (1971) *Ecrits logiques et philosophiques*, Paris, Seuil.
- Gonseth, F. (1990) *Le problème de la connaissance en philosophie ouverte*, Lausanne, L'âge d'homme.
- Kripke, S. (1972) *La logique des noms propres*, Paris, Minuit.
- Leibniz (1970) *La Monadologie*, Paris, Delagrave (éd. E. Boutroux).
- Putnam (1994) article Putnam, in Guttenplan, S. (éd.) *A Companion to the Philosophy of Mind*, Oxford, Blackwell, pp. 507-513. [tr. fr. à paraître sous le titre *Aller et retour au pays du fonctionnalisme*].
- Rastier F. (1991) *Sémantique et recherches cognitives*, Paris, PUF.
- Récanati F. (1991) Du langage à la pensée, *Actes du colloque Sciences de la cognition*, Paris, MRT, pp. 137-141.

\*\*\*

Jean-Michel ROY. — *Qu'est-ce qu'une explication fonctionnelle ?*

Chacune des trois questions posées soulève à mon sens sous un angle différent mais complémentaire des deux autres le problème de la définition du fonctionnalisme dans le contexte de ce qu'il est aujourd'hui convenu d'appeler les sciences de l'esprit. Or ainsi entendues, ces questions peuvent se reformuler de la manière suivante : 1°) Ce que l'on désigne couramment par fonctionnalisme, à savoir la théorie computationnelle classique de l'esprit, ne représente-t-il en fait qu'une simple espèce ? 2°) Le cas échéant, quelles autres espèces de fonctionnalisme sont possibles ? 3°) En outre, dans quelle mesure les théories de l'esprit existantes qui se présentent aujourd'hui comme des alternatives au fonctionnalisme classique d'une part en diffèrent-elles effectivement d'une part, et d'autre part incarnent-elles précisément des formes différentes de fonctionnalisme ?

Je suis pour ma part enclin à considérer qu'il est possible d'apporter une réponse positive à la première d'entre elles, reconnaissant du fait même les deux autres comme pertinentes. Mais plus encore, il me paraît qu'une telle réponse est d'une parfaite évidence et que c'est en fait la première question elle-même qui dans une large mesure manque de

pertinence. Il me semble en effet illégitime de supposer que le fonctionnalisme computationnel classique se soit jamais présenté comme l'unique forme possible du fonctionnalisme, même s'il n'a sans doute pas marqué avec une suffisante netteté son inscription au sein d'une doctrine plus générale, ni analysé avec une précision satisfaisante la nature tant du genre auquel il revendiquait ainsi l'appartenance, que de la différence spécifique qui était censée être la sienne. Et aujourd'hui encore les études critiques sur le fonctionnalisme mental se bornent pour la plupart à des remarques indicatives à cet égard, bien qu'il s'agisse là des questions épistémologiques les plus fondamentales qu'il soulève. Cette pauvreté épistémologique de la réflexion sur le fonctionnalisme contemporain dans le domaine des sciences de l'esprit doit d'ailleurs être contrastée avec l'intérêt suscité par le fonctionnalisme sociologique et anthropologique chez certains éminents philosophes de la science à la charnière des années 50 et 60<sup>1</sup>. Malgré le désaccord partiel que je viens de souligner avec l'affirmation qui leur sert de prémisse, il me paraît donc que les questions précédentes renvoient à une interrogation d'une grande importance et d'une grande justesse que je préfère exprimer pour ma part sous la forme peut-être plus systématique suivante : A) Qu'est-ce qui caractérise une explication de type fonctionnel en général ? Dans quelle mesure mérite-t-elle d'ailleurs le qualificatif d'explication ? B) Qu'est-ce qui distingue l'explication fonctionnelle contemporaine dans le domaine des sciences de l'esprit des formes qu'elle peut prendre ou avoir prises dans ceux d'autres disciplines telles que les sciences sociales, ou encore que la physiologie ? C) Quelle typologie peut-on dresser des explications fonctionnelles dans le domaine du mental ? Comment en particulier peut-on y situer le fonctionnalisme computationnel classique ?

Mais du fait même ces questions pointent vers une lacune substantielle lacune de la littérature contemporaine sur le fonctionnalisme mental qu'il ne saurait être question de combler en un espace de réflexion aussi exigü. Je me contenterai par voie de conséquence, en prenant pour point de départ une justification de la critique qui vient d'être formulée à l'encontre de la manière dont se présente le fonctionnalisme computationnel, d'essayer de dégager quelques éléments d'une définition générale de l'explication fonctionnelle (question A) qui puisse servir de critère de base à la fois pour une exploration conceptuelle de ses formes possibles (question B), dans laquelle je n'entrerai pas, et pour une évaluation des théories disponibles, que je serai également contraint de négliger entièrement (question C).

---

<sup>1</sup> Cf. notamment Carl Hempel, « The Logic of Functional Analysis », *Aspects of Scientific Explanations*, 1965, et E. Nagel, *The Structure of Science*, 1960.



Il est incontestable que certaines des formulations originelles de l'hypothèse fonctionnaliste dans le domaine des sciences de l'esprit ont négligé d'opérer une distinction soignée entre fonctionnalisme et fonctionnalisme computationnel à la Turing. On pourrait à l'appui invoquer par exemple le célèbre article de Putnam « La nature des états mentaux »<sup>1</sup>, où celui-ci assimile d'emblée la notion d'organisation fonctionnelle d'un organisme à celle de sa description en termes de machines de Turing, et le concept d'état fonctionnel à celui d'un état d'une telle machine. Toutefois dans des articles plus anciens, tel que « Minds and Machines », ces couples de notions se trouvent en fait clairement distingués et le rapport qui les unit est présenté comme étant de nature purement analogique ; Putnam se contente en effet alors de suggérer que la caractérisation mentaliste de l'étiologie du comportement d'un organisme est, avec la caractérisation matérialiste de cette étiologie, dans la même relation d'abstraction que la caractérisation du fonctionnement d'un ordinateur en termes de machine de Turing entretient avec sa caractérisation en termes purement physiques. Le fait qu'il ait, dans un deuxième temps, proposé d'identifier l'un avec l'autre les deux niveaux abstraits d'explication ne doit pas masquer qu'il propose dès le départ une définition générale de l'explication fonctionnelle comme explication causale abstraite, ou encore comme schéma d'explication causale. Il convient d'ailleurs d'ajouter que l'explication fonctionnelle ainsi conçue n'est pas selon Putnam intrinsèquement liée à une ontologie matérialiste, mais compatible au contraire avec une position dualiste<sup>2</sup>. Cette définition générale se trouva d'ailleurs très vite mise en lumière par son élève Jerry Fodor qui, au chapitre III de *L'explication en psychologie*<sup>3</sup>, défend l'idée qu'une explication psychologique, considérée dans une perspective fonctionnaliste, formule une hypothèse explicative sur les processus causaux cérébraux responsables du comportement d'un organisme, mais, à la différence de l'explication neurobiologique qui pour cette raison doit venir ensuite la compléter, ne précise que leurs propriétés causales et pas leurs propriétés physiques. En bref, une explication fonctionnelle portant sur le mental est définie comme un niveau abstrait de l'explication causale d'un processus physique en ce sens que l'on n'y spécifie pas ce que sont les éléments qui dans ce processus ont telle ou telle propriété causale. Ce n'est que dans un second moment de son entreprise<sup>4</sup> que Fodor introduit, à

---

<sup>1</sup> Cf. La théorie computationnelle de l'esprit, *Etudes Philosophiques*, 1992, n° 3.

<sup>2</sup> « En particulier, l'hypothèse de l'état fonctionnel n'est pas incompatible avec le dualisme ! », *op. cit.*, p 330.

<sup>3</sup> Seghers 1972, trad. fr. de Y. et G. Noizet.

<sup>4</sup> Cf. *The Language of Thought*, 1975.

côté du niveau mentaliste conçu fonctionnellement, un niveau computationnel (classique) d'explication qui à son tour est conçu comme une spécification du niveau mental, et de ce fait comme étant lui-même fonctionnel au sens qui vient d'être précisé.

Mais qu'une explication fonctionnelle de type computationnel puisse être véritablement assimilée à une forme d'explication fonctionnelle de type causal demande à mon sens démonstration. Les computationnalistes, à ma connaissance, ne se sont jamais inquiétés de la fournir. La critique du fonctionnalisme computationnel proposée par E. Sober <sup>1</sup> repose d'ailleurs de manière cruciale sur la supposition que le fonctionnalisme computationnel n'est pas considéré comme une forme du fonctionnalisme causal, qu'il appelle quant à lui fonctionnalisme téléologique, et que les difficultés qui affectent le premier peuvent donc ne pas atteindre le second. Mais cette supposition me paraît tout à fait inexacte, de même d'ailleurs que me paraît inadéquat le choix du terme de téléologique.

Dans son excellente quoique trop brève tentative de définition du fonctionnalisme, N. Block <sup>2</sup> inscrit d'ailleurs lui aussi sans ambiguïté le fonctionnalisme computationnel au sein du genre de ce qu'il appelle l'analyse fonctionnelle et qui pour l'essentiel correspond à ce qui vient d'être décrit sous le titre de fonctionnalisme causal. La distinction qu'il opère toutefois entre le genre de l'explication fonctionnelle ainsi conçue et le fonctionnalisme métaphysique comme théorie de la nature des états mentaux me semble en très large partie artificielle, puisque l'explication fonctionnelle causale se caractérise fondamentalement par un mode d'individuation causal des états mentaux en tant qu'éléments des processus mentaux. En outre, d'une part il néglige le fait que sa définition générique du fonctionnalisme ne puisse pas prendre en compte les formes de fonctionnalisme véritablement téléologique qui sont apparues à la suite du fonctionnalisme computationnel, et d'autre part il renonce a priori à confronter le fonctionnalisme des sciences de l'esprit avec celui d'autres disciplines. Deux omissions qui se rejoignent puisque le fonctionnalisme anthropologique et sociologique, à en croire du moins les analyses de Nagel et de Hempel, se caractérise par l'appel non pas seulement à l'idée de cause efficiente mais aussi de fin. Or, sauf à pouvoir montrer que toute fonction peut être réduite à une cause efficiente, il semble qu'on ne puisse donc admettre comme définition générique du fonctionnalisme la définition causale préalablement dégagée. Mais du même coup il apparaît

---

<sup>1</sup> « Panglossian Functionalism and the Philosophy of Mind », *Synthese* 64, n° 2, 1985, pp.165-93.

<sup>2</sup> « What is Functionalism? », *Readings in the Philosophy of Psychology*, Harvard University Press, 1978.

aussi maintenant clairement qu'il doit y avoir en principe autant de formes d'explication fonctionnelle qu'il y a de types irréductibles de fonction, et que la définition générale du fonctionnalisme est en fait suspendue à celle de la notion de fonction. A la vérité, c'est moins le concept de fonction que celui de fonctionnement — la manière dont quelque chose fonctionne — qui me semble pertinent pour caractériser l'objet d'une explication fonctionnelle. De telle manière qu'il est tentant de suggérer qu'une explication est fonctionnelle au sens le plus large du terme quand elle explique le fonctionnement de quelque chose, et qu'il en existe autant de types qu'il existe de genres de fonctionnement. Mais une telle définition n'est-elle pas alors inversement trop indéterminée ? Toute explication n'est-elle pas en effet explication de la manière dont quelque chose fonctionne, et à ce titre une explication fonctionnelle ? J'en conviens pour ma part volontiers et propose par conséquent de retenir en outre dans le concept spécifique d'explication fonctionnelle l'idée que cette explication est abstraite, au sens où l'entend Fodor en 1968 dans le cas particulier de l'explication mentale fonctionnelle de type causal. En d'autres termes une explication serait fonctionnelle au sens spécifique du terme quand elle explique la manière dont quelque chose fonctionne mais sans éclairer la dimension physique de ce fonctionnement.

Jean-M. ROY  
 UFR de Philosophie  
 Université Michel de Montaigne  
 Domaine Universitaire, 33405 Talence Cedex

\*\*\*

Jürgen SCHRÖDER. — *Functionalism, machines, and propositional attitudes*

The original aim of functionalism has been to answer the question what the nature of mental states is, i.e. what the properties are in virtue of which something is a mental state of a certain kind. The fundamental idea of functionalism is that it is its causal role that makes a physical state a mental state. This idea can be specified in different ways according to what theory is taken to determine the causal roles. Functionalism depends in all cases on such a theory but doesn't by itself decide between any of various possible candidate theories. It is possible, for example, to take as the "theory" all the "platitudes" of common sense concerning the causal relations between stimuli, behaviour and mental states and to determine the various causal roles by these platitudes, as David Lewis did once suggest.

Or you may turn to psychological theories and arrive at a different characterization of the causal roles.

Neither choice implies a reference to machines, however. How then are machines wedded to functionalism? Machine versions of functionalism are a response to the ever present risk that functional explanations might be pseudo explanations. One form of pseudo explanation consists in the fact that the causally relevant property of a state  $s_1$  for a state  $s_2$  is described as "the capacity of  $s_1$  to bring about  $s_2$ ". This is the same kind of explanation as that of explaining the effect of morphine by its soporific power. Another form of pseudo explanation refers to functional descriptions that can be realized by nothing, e. g. a function that maps any possible question onto its answer (cf. Fodor, 1981, p. 12).

In order to exclude such empty or incoherent explanations a functional description can be couched in the vocabulary of Turing machines. Once this has been done one can be sure that if one has a TM program which is input-output equivalent to a certain cognitive capacity one has a *possible* explanation of this capacity and not a pseudo explanation. If the explanation is to be a good one further criteria have to be satisfied. There has to be something stronger than merely input-output equivalence. The notion of such a "strong" equivalence has been elaborated by Pylyshyn (1984). One of Pylyshyn's conclusions was that in order to establish the strong equivalence between mental and machine program the functional architecture of the brain must be known. Then a program that is executed by a virtual machine that has the same functional architecture can be compared according to certain criteria like complexity equivalence (does the virtual machine take the same number of steps in order to compute a certain function as the brain?), etc. with the mental program of the brain.

The functional architecture of Turing machines is, however, not the one in virtue of which one might achieve a strong equivalence between computational models and mental processes for reasons that Block and Fodor have given (Block & Fodor, 1972) and because of constraints such as the 100 step constraint (cf. Feldman & Ballard, 1982).

Functional explanations can be made precise by expressing them in the vocabulary of a formal system thereby avoiding pseudo explanations. But additional conditions must be met in order for a certain formal system to be adequate.

The question whether every cognitive architecture must at least have a symbolic part, that is finally whether complex representations must have parts that correspond to their constituents is a more difficult one. Properties of cognitive capacities as productivity, systematicity, and

compositionality seem to require a symbolic architecture (Fodor & Pylyshyn, 1988). On the other hand there are a number of attempts by connectionists to develop networks that display a restricted form of productivity and systematicity (e.g. Chalmers, 1990; Elman, 1989; Niklasson & van Gelder, 1994; Pollack, 1991, Smolensky, 1991; St. John & McClelland, 1991). The problem concerning systematicity consists in the fact that the architecture of connectionist systems doesn't guarantee systematicity and that there is no general proposal for the reconstruction of the modes of propositional attitudes. The discussion centers around an adequacy condition for cognitive theories, namely that it specify the constitutive bases for intentional capacities (McLaughlin, 1993). These constitutive bases consist in the capacity to token concept representations, the capacity to jointly exercise these concepts, and the capacity to display the result, the propositional representation, in a certain intentional mode. If we restrict ourselves to perceptual capacities natural selection could guarantee systematicity and Smolensky's tensor product representations could be constitutive bases for concepts and their joint exercise. The mode "perceptual belief" would be constituted by the representations being tokened in networks dedicated to perception (Schröder, 1995).

But what is functionalist about a connectionist network? If this question is to be understood as referring to what is functionalist about a connectionist account of propositional attitudes the answer would be: nothing. For, in classical architectures the only thing that gets a functionalist interpretation is the modes of propositional attitudes (barring conceptual role semantics) and in connectionism there is no general account of modes. So we don't know yet what the functional element would be in connectionist systems that have "realist" propositional attitudes. If, however, the question refers to cognitive capacities the functionalist element would consist in the analysis of a cognitive capacity into a configuration of subcapacities such that each subcapacity is realized by a connectionist network. It is remarkable that models of cognitive psychology that try to give such analyses don't seldom make reference to cognitive states such as propositional attitudes and that the question "what does a belief consist in?" is not a question for which you might get easily an answer out of a psychological model.

Jürgen SCHRÖDER  
CREA  
1, rue Descartes  
75005 Paris

**BIBLIOGRAPHY**

- Block, N. & Fodor, J., What psychological states are not, *Philosophical Review*, 81, 1972, 159-82.
- Chalmers, D., Syntactic transformations on distributed representations. *Connection Science*, 2, 1990, 53-62.
- Elman, J.C. Representation and structure in connectionist models. CRL Technical Report 8903.
- Feldman, J. & Ballard, D., Connectionist models and their properties. *Cognitive Science*, 6, 1982, 205-54.
- Fodor, J. *Representations*, Brighton: The Harvester Press, 1981
- Fodor, J. & Pylyshyn, Z. Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3-71.
- McLaughlin, B., Systematicity, Conceptual Truth, and Evolution, in C. Hookway & D. Peterson (eds.), *Philosophy and Cognitive Science*, Royal Institute of Philosophy, supplement, No. 34, 1993, 217-34.
- Niklasson, L. & van Gelder, T., On Being Systematically Connectionist, *Mind & Language*, 9, 1994, 288-302.
- Pollack, J.B., Recursive distributed representations, in G. Hinton (ed.) *Connectionist Symbol Processing*, Cambridge: MIT Press, 1991.
- Pylyshyn, Z., *Computation and Cognition*, Cambridge: MIT Press, 1984.
- Schröder, J., Systematicity, Compositionality, and the Notion of Symbolic Representations, *Proceedings of the Second World Conference on the Fundamentals of Artificial Intelligence*, 1995, 379-92.
- Smolensky, P., Tensor product variable binding and the representation of symbolic structures in connectionist systems, in: G. Hinton (ed.) *Connectionist Symbol Processing*, Cambridge: MIT Press, 1991.
- St. John, M.F. & McClelland, J.L., Learning and applying contextual constraints in sentence comprehension in: G. Hinton (ed.) *Connectionist Symbol Processing*, Cambridge: MIT Press, 1991.

\*\*\*

Yves-Marie VISETTI. — *Fonctionnalismes 1996*

Nous partirons d'un premier repérage de l'espace des fonctionnalismes, celui que nous propose Elisabeth Pacherie dans son appel à contribution<sup>1</sup>.

---

<sup>1</sup> Le texte présenté ici devait être ma contribution au débat sur les fonctionnalismes, auquel j'étais comme tout un chacun appelé à participer sous la forme libre, mais réduite, d'un *jalon*. Mais il a proliféré en chemin et fini par ressembler à un article véritable, sans pour autant satisfaire à toutes les normes académiques en usage. Je tente d'y présenter, avec la précision et dans le style que je crois affines à mon propos,

Elle suggère en effet trois dimensions principales, présumées caractéristiques de cette famille de paradigmes cognitifs. Tout fonctionnalisme comprendrait ainsi – ou du moins laisserait supposer – un triple choix constitutif : choix d'un type de “machine”, choix d'un type de “formes” aptes à s'y inscrire, et choix enfin d'un type de “valeurs” susceptibles d'être attribuées à ces formes. Nous adopterons cette grille d'analyse, d'autant plus qu'elle nous permettra de mettre en avant la notion de forme, plutôt que celle de fonction. Le premier fonctionnalisme, celui du jeune Putnam, de Fodor ou Block, insistait plutôt sur l'irréductibilité des descriptions fonctionnelles. Mais les fonctionnalismes plus récents se présentent au contraire comme des philosophies de la naturalisation : il leur est donc vital de faire reposer les fonctions sur des formes naturalisables, de préciser davantage quelles formes dans la nature peuvent ou non être recrutées au service d'une description fonctionnelle. Quitte à déplacer quelque peu la perspective, je souhaite par conséquent les reconstruire aussi comme des philosophies de la forme, ou tout au moins cerner ce qui s'apparente en elles à une telle philosophie, car c'est de ce point de vue, je crois, qu'on peut le mieux analyser les interminables résurgences du fonctionnalisme.

1. *La machine.* — Selon ce point de vue, par conséquent, le premier geste – assurément le premier dans l'ordre de la fondation, si ce n'est dans l'ordre de la découverte – le premier geste des fonctionnalismes serait de se donner une machine, que l'on conçoit ici comme une étendue matérielle “formatée” d'une façon bien déterminée. On entend par là qu'une machine est, de par sa seule structure physique, le siège de processus effectifs respectant un certain “format”, c'est-à-dire un type de distinctions physiques élémentaires dont le principe d'assignation nous est connu, du moins localement. Toute l'effectivité processuelle du dispositif doit ainsi pouvoir se décrire au niveau de ce même format, c'est-à-dire uniquement à partir du champ de distinctions permises, et sans revenir à toute la complexité physique sous-jacente. Et c'est en cela précisément que consiste toute la finalité de telles machines : elles ne sont pas tant définies par leur asservissement à une fin particulière, mais d'abord par le

---

ma perception globale des fonctionnalismes, dans leur variété actuelle. Je leur oppose d'autres postulats, d'autres façons de parler et de penser, qui ne sont évidemment pas seulement les miens. Je dis enfin le sentiment profondément critique que m'inspire leur relation aux disciplines de la cognition et à tout le reste de la philosophie. Il en découle un texte dont le genre déroutera peut-être le lecteur, ce dont je crois plus juste de l'avertir. Ce texte n'en est pas moins construit, et je l'ai divisé en sept sections intitulées : *la machine, les formes, les valeurs, quelques fonctionnalismes, postulats et exercices, pour une notion de champ, et polémiques.*

fait qu'elles abritent et maintiennent dans leur clôture une certaine factorisation assignable du réel physique. La finalité première est donc la stabilisation d'un certain formatage physique, qui ne serait, comme tel, astreint à aucune fin particulière donnée à l'avance. A la base des fonctionnalismes, on trouve donc un statut paradoxal des machines – instances d'une finalité sans fins préassignées – qui les distingue radicalement des machines qui les ont précédées. Il s'agit de la finalité que nous attribuons à un certain format physique spécifique : celle de déployer, au service de fins encore à définir, une objectivité interne dont l'exacte détermination, tant théorique qu'empirique, s'oppose à la diversité impensable et incontrôlable des configurations matérielles environnantes.

Tout observateur conscient de la factorisation réalisée dans le formatage peut alors considérer ces étendues matérielles comme des substrats porteurs de *marques* modifiables. Ces marques peuvent d'ailleurs présenter des types de stabilité différents d'une machine à l'autre ; mais c'est toujours leur temporalité propre d'apparition ou de disparition, leur spatialité caractéristique, qui constituent la temporalité et la spatialité de référence de la machine. Ainsi donc “l'encre ” dont ces marques sont “écrites ” se dissout plus ou moins vite dans son substrat, tandis qu'une machine n'est rien de plus que le support d'un certain type “d'écritures ” plus ou moins réactives. Autrement dit encore, les machines s'identifient à d'énormes diagrammes animés selon une loi propre à leur support, elles déploient une diagrammatique temporelle et systématique. Il suffit alors de savoir “lire ” et “écrire ” ces “diagrammes ” pour disposer d'un nouveau principe d'intelligibilité et de contrôle des événements survenant à l'intérieur de la clôture machinique. La machine de Turing et sa réalisation finie, l'ordinateur à la Von Neuman, est évidemment le principal cas de figure de cette situation ; ses marques de référence sont, comme il est bien connu, généralement notées 0 et 1 ; son temps de référence est discret, etc. Mais les calculateurs analogiques aujourd'hui disparus, ou les actuels réseaux de neurones formels (matérialisés ou non) en sont d'autres instances possibles ; leur marquage spécifique pourrait s'identifier aux résultats d'une certaine mesure physique, et donc présenter d'emblée la diversité d'un sous-ensemble des nombres réels ; leur temps de référence pourrait être continu ou discret, etc.

Jusque là pas de problèmes, car il s'agissait d'artefacts, réels ou virtuels. Mais l'idée fonctionnaliste est précisément d'étendre cette même approche aux êtres vivants et/ou pensants, de façon à *constituer comme machines* les psychologies humaines ou les cognitions animales. S'il est donc un “matérialisme ” conséquent, qui prétend valoir un jour empiriquement, un fonctionnalisme strict ne postulera que des



distinctions objectives toutes inscriptibles dans le marquage d'une machine bien déterminée. Bien plus encore, il posera que l'être cognitif n'est tel que parce qu'il est lui-même machine, instance de machine. La détermination *matérielle* précise des formatages demeure évidemment un objet de spéculation ; toutefois, il apparaît généralement que, s'il s'agit d'êtres vivants, les cerveaux sont concernés, et eux seulement. Quant aux déterminations *formelles* de ces mêmes formatages, elles se rangent principalement sous deux catégories : machines de Turing d'une part, et systèmes dynamiques d'autre part (les réseaux connexionnistes en sont un cas particulier).

L'appel fonctionnaliste aux machines ne signifie donc pas qu'on aurait déterminé sur cette base un accès systématique à un vaste ensemble expérimental, biologique, psychologique, ou autre. Il signifie simplement qu'on a réifié certaines heuristiques interprétatives, et fait un usage métaphysique de certains principes de réflexion, en leur donnant le statut de concepts déterminant des objets accessibles dans une expérience. Les êtres vivants deviennent alors les hypostases d'un postulat métaphysique, le formatage, qui prescrit d'avance la factorisation des phénomènes dans un certain système de marques, alors même que ces marques demeurent non déterminées dans leur matérialité. Pour ces raisons, les fonctionnalismes se disent souvent matérialistes et non réductionnistes. Ainsi l'agencement singulier de la matière (le substrat physique de la machine) demeure totalement indifférent aussi longtemps qu'aucune différence ne se marque au niveau du formatage (qui présuppose cependant une matière absolument conforme à sa loi)<sup>1</sup>.

---

<sup>1</sup> Le dualisme latent des fonctionnalismes s'exprime dans cette indifférence. Plus encore, l'identité occasionnelle ou *token identity*, qui relierait les états physiques aux présumés "états mentaux" qu'ils "instancient", interdit d'identifier un état mental à un type physique déterminé une fois pour toutes. Des êtres de structures physiques différentes, ou le même être considéré à différents moments, peuvent donc se trouver dans le même état mental sans pour autant être dans le même état physique. Cela semble aller contre l'idée de format machinique, qui suppose une assignation de distinctions physiques bien déterminée à tout moment. Cependant, les fonctionnalismes affirment que la distinction entre états mentaux, et l'identité à soi d'un état mental, sont des notions ontologiquement fondées ; ils affirment d'autre part que deux états mentaux différents A et B sont nécessairement différents physiquement, au sens où l'on doit pouvoir, à un instant déterminé et pour un être déterminé, distinguer physiquement A et B. Je ne vois donc pas comment ils pourraient se passer de postuler un ensemble déterminé de distinctions physiques données en présence à chaque instant et relatives à chaque être, en contrepartie de la permanence des distinctions identitaires entre états mentaux (qui se font en quelque sorte concurrence pour advenir à ce même instant). Ce serait bien là le double engagement des fonctionnalismes : pour une ontologie d'états mentaux parfaitement individués, d'une

On objectera peut-être que la description que je viens de donner ne vaut que pour certains types de fonctionnalismes, actualisant dans un cadre “matérialiste” le projet d'une psychologie rationnelle (on peut penser à Fodor), et vaut déjà moins pour d'autres fonctionnalismes, de tonalité plus empiriste (les biofonctionnalismes, par exemple). J'admets la valeur de cette objection éventuelle, mais je n'y souscris pas. Un problème fondamental des fonctionnalismes semble en effet de trouver un passage systématique entre la notion de structure matérielle et celle de fonction, en imaginant toutefois que *les structures matérielles sont déjà données et individuées dès le départ*, si bien que les conditions de leur morphogenèse seraient d'un ordre radicalement différent de celles du “fonctionnement” de plein exercice, qui pourrait et devrait dès lors être compris comme réseau de relations abstraites. Bien que le concept de machine ne soit pas nécessairement mobilisé de façon explicite à cet endroit, il m'apparaît que se donner dès le départ des “structures” matérielles individuées implique précisément de postuler une factorisation de la complexité biologique en un format biologique dont l'auto-reproduction ne poserait plus problème, sur lequel il n'y aurait donc plus à revenir pour comprendre la relation du vivant à son milieu. Tout à l'opposé, par conséquent, d'un fonctionnement ressaisi dans le cadre d'une ontogenèse allant de la conception à la mort.

Quels sont donc, en résumé, les caractères les plus courants d'une “mise au format” fonctionnaliste ?

– Il s'agit d'une véritable disqualification de la complexité physique et/ou biologique, qui joue comme préalable essentiel à la constitution du domaine cognitif en tant que tel ; tout le contraire, donc, d'un geste de modélisation ordinaire en sciences de la nature, qui consiste en une schématisation respectueuse de la diversité empirique, peut-être décisive, qu'elle laisse échapper.

---

part, pour un ensemble sous-jacent de distinctions physiques tout aussi déterminées (à chaque instant, pour chaque être), d'autre part. Je sais bien que la “matière” des fonctionnalismes n'est qu'une empirie métaphysique (en attendant mieux) dont les principes de détermination ne sont pas soumis à une juridiction critique, mais je pense être déjà arrivé à l'extrême limite des concessions. Au-delà, il faudra jouer sans moi. Cela étant dit, il reste admissible que le format machinique (en tant que système global de distinctions physiques) puisse, d'instant en instant, varier le mode matériel de son inscription, tout en restant fidèle au type abstrait de ses marques ; la machine, en quelque sorte, investirait sa propre matière d'une façon perpétuellement décalée, tout en se préservant dans ce décalage. Ce n'est pas totalement impossible à imaginer, mais il me semble dans ce cas que seuls les fonctionnalismes dynamiques (cf. infra), dans la mesure où ils s'autorisent d'une théorie de l'émergence, pourraient combiner l'invariance systémique avec la dérive de son inscription matérielle.

– L'unité interne du vivant est celle, homogène, du formatage machinique ; combiner plusieurs machines rend peut-être cette unité plus diverse, mais ne fait pas sortir du formatage ; car un fonctionnalisme cohérent sera tenu d'assurer les transferts d'un format à l'autre (postuler un milieu interstitiel non formaté serait déjà s'éloigner du fonctionnalisme) : et on obtiendra alors une seule machine distribuée en modules.

– Le formatage interne est acquis une fois pour toutes à l'échelle d'un individu ; il ne se renouvelle pas, et ne dépend pas, serait-ce pour sa reproduction, des fins particulières poursuivies par le dispositif, de la qualité générale de la relation à l'environnement, de l'histoire, etc. ; toutefois, les biofonctionnalisments diffèrent à cet égard, en raison de leur perspective transgénérationnelle : il peut donc y avoir mutation du formatage d'une génération à l'autre, et la machine se reproduit plus ou moins bien, peut-on penser, selon "l'avantage" relatif conféré par son formatage propre ; cependant, il s'agit là d'un phénomène en tout ou rien, et comme nous l'avons déjà dit morphogénèse et fonctionnement restent séparés ; si bien que dans la plus pure tradition néo-darwinienne, l'individuation des formats de machines est supposée acquise, avant que la sélection, comme son nom l'indique, soit appelée à choisir dans leur répertoire.

– La finitude de la machine n'entre pour rien dans la définition du format ; il peut être conçu à partir d'un substrat idéalisé comme infini ; passant au réel, il subit une simple troncature, qui ne transforme pas le principe de marquage (ainsi l'ordinateur et la machine de Turing) ; il s'agit donc d'une finitude simplement limitante, et non constituante.

– La machine est matériellement ou énergétiquement ouverte sur l'extérieur ; en conséquence de ce qui précède, le flux entrant ne peut intervenir que s'il est *transducté*, c'est-à-dire mis au format ; au-delà, sa diversité propre est complètement ignorée.

2. *Les formes.* — Le second geste des fonctionnalisments consiste à élaborer à partir du format de base, pour reconnaître ou imprimer dans la machine un certain éventail de formes avec leur gamme de transformations possibles. L'enjeu est bien sûr de compenser par cette diversité nouvelle l'homogénéité et l'inaltérabilité du format englobant, et donc de rendre possible un couplage efficace à l'extériorité. L'accroissement de la diversité ne repose donc pas sur une diversification du format ; elle consiste seulement en un raffinement de la perception de l'observateur qui scrute le flot physiquement formaté de la matière machinique ; il distingue ainsi et catégorise des événements de grain

variable<sup>1</sup>, selon des modalités toujours plus nombreuses et stables (impliquant par exemple segmentation et typage).

Le fait capital, à ce point, est encore une fois que la détermination de ces formes et de ces transformations peut se faire, dans une très large mesure, sans aucunement tenir compte de la matérialité du dispositif. Ce n'est pas que les considérations de taille, de vitesse, ou de permanence des bases physiques du formatage soient toujours ignorées : bien au contraire, de nombreuses approches fonctionnalistes (de la mémoire, par exemple) mettent en avant des questions d'embouteillage, de dépassement de capacité, de lésions, etc. De même une certaine conception démiurgique de l'Intelligence Artificielle a pu mettre l'accent sur les caractéristiques singulières des ordinateurs, pour mieux soutenir son pari métaphysique. Mais l'essentiel est ailleurs : il consiste à mettre une fois pour toutes les *répertoires* de base et les *systematicités* des formes à l'abri de toute détermination matérielle, intérieure ou extérieure, qui ne se factoriserait pas dans le format prévu : les formes changent donc, mais dans le cadre prévu par la loi du formatage.

Les systèmes adaptatifs, ou à apprentissage, ne dérogent d'ailleurs pas à ce principe : changer de représentations symboliques, modifier les poids d'un réseau connexionniste, faire bifurquer un système dynamique en manipulant un paramètre de contrôle, ne transforment pas le format de base (représenté par la bande de la machine de Turing, ou l'espace d'états du réseau), mais seulement le système des formes qui s'y inscrivent. Toutefois, on peut dire que les effets obtenus évoquent parfois le recrutement d'une nouvelle distinction de base qui imprimerait, d'un seul coup, ses effets différenciateurs à l'intérieur de toutes les formes en jeu. D'ailleurs, nous pouvons d'autant mieux simuler un tel événement que nous pouvons recommencer avec les formes le même jeu d'abstraction qui nous avait mené de la matière jusqu'à la machine : nous pouvons choisir un certain répertoire de formes catégorisées en "types", auquel nous ferons jouer le rôle d'un format de base pour une nouvelle machine, dont nous pourrions dire qu'elle est simulée par la précédente, etc. On aura ainsi établi une sorte d'équivalence entre les concepts de machines et de formes, que la notion d'architecture résume assez bien. Et si par ailleurs chaque machine semble bien avoir ses formes favorites (relativement à la perception de l'observateur), on sait qu'elle peut, dans une certaine

---

<sup>1</sup> On connaît, de ce point de vue, la limite décisive des fonctionnalismes logiques : pour eux, la différence des niveaux ou grains de description se ramène toujours à une combinatoire finie (celle de la compilation, par exemple). Si bien qu'aucun niveau d'appréhension des formes ne semble pouvoir induire de décalage ontologique par rapport aux autres.

mesure, donner aussi à voir les formes plus couramment associées à ses concurrentes : ainsi des machines de Turing, qui supportent des approximations discrètes des systèmes dynamiques, ou des réseaux connexionnistes, qui à l'inverse pourraient approcher une configuration d'ordinateur classique<sup>1</sup>.

Y a-t-il alors indépendance de la formalité des formes relativement à la finitude du dispositif, comme c'était toujours le cas pour le format de base de la machine ? Ici les fonctionnalismes divergent. Les formes logiques ne dépendent pas constitutivement d'une finitude englobante, si ce n'est pour des questions de taille (donc un simple principe de limitation) ; et pour ce qui est du temps, elles imposent aux processus qui les transforment une structure discrète. De leur côté les formes des fonctionnalismes dynamiques peuvent subir *la pression constituante*, topologique, géométrique, et temporelle, de multiples contraintes globales dont dépend leur individuation. Il n'y a plus alors de distinctions nettes entre structures et processus : car les formes ne sont que les segments d'un processus qui ne s'interrompt jamais (dans les limites autorisées par le format machinique sous-jacent). C'est la structure globale du temps significatif et de l'espace disponible qui détermine alors ce qui peut advenir localement et provisoirement comme forme : là réside l'atout maître des fonctionnalismes dynamiques, en même temps que leur limite (à savoir la méconnaissance non problématisée des conditions matérielles de ces structures spatio-temporelles globales).

Reste dans tous les cas à varier les formes, leur grammaire ou leur répertoire, de façon à soutenir le jeu de certaines valeurs ou fins qui dessinent l'univers du système "survenant" à la machine et "fonctionnant" sur elle.

3. *Les valeurs.* — Attribuées aux formes par l'observateur, elles constituent l'intelligibilité ultime du système, la preuve qu'il fallait en passer par le dispositif machine + formes pour comprendre comment le système respecte en définitive les différences que ces valeurs qualifient. Toute la question est donc là : comment attribuer la valeur, et comment joue-t-elle implicitement dans la dynamique des formes ? On imagine bien qu'un fonctionnalisme strict, pour ne pas dire intégriste, pourrait faire l'impasse totale sur l'extériorité non proximale, et définir la valeur d'une forme comme son rôle purement différentiel relativement aux autres formes et à la transduction sensori-motrice. Mais il est décidément

---

<sup>1</sup> Les fonctionnalismes peuvent donc admettre l'hétérogénéité apparente des formes, dès le moment où elle se laisse en fin de compte réduire au format d'une machine unique.

difficile de comprendre le sens d'une conduite humaine ou animale si l'on fait totalement abstraction de son monde. Renoncer au monde serait aussi renoncer au sens, même si le sens n'est pas qu'effets de sens présents dans le monde. C'est donc ici que le tableau commence à présenter quelque confusion, avec toutefois un invariant d'un fonctionnalisme à l'autre : le sens en tant que rapport au monde semble toujours constitué sur le mode de la représentation, ou du moins c'est la capacité de représenter, en tant qu'accès à la vérité ou possibilité de la survie, qui concentre la valeur et mobilise le système à son service.

Ainsi donc, même si nous supposons que “représenter” est une notion qui ne pose pas de problème, nous voyons clairement que les fonctionnalismes, s'ils veulent gagner l'extériorité qu'ils ont commencé par écarter, doivent résoudre ou contourner deux problèmes :

– si les formes ont été constituées indépendamment de cette relation présumée entre le système et son monde (la capacité de le représenter), il faut expliquer comment elles pourraient contribuer à établir cette relation dans l'après-coup de leur constitution ;

– si nous partons d'une pré-compréhension intuitive de “représenter” comme d'une relation désintéressée au monde, il faut expliquer en quoi le fait de représenter peut être, en tant que tel, l'instrument principal d'un intérêt pour le monde, d'une orientation ou directionnalité des conduites, bref d'une intentionnalité.

4. *Quelques fonctionnalismes.* — Cela étant posé, voici pêle-mêle et sans trop d'égards pour les conventions académiques, quelques éléments tombés du tableau des fonctionnalismes.

a) *Fonctionnalisme logique.* — Une recette bien connue à présent. Prendre une machine de Turing, inscrire ou lire des formes catégorisées en types logiques, et des processus matérialisant des procédures formelles commandées par la syntaxe. S'appuyer ensuite sur la théorie logique des modèles, pour faire référence à une extériorité virtuelle qui n'a en définitive pas d'autre nature que celle d'une idéalité logique. Constituance et donc compositionnalité. Internaliser de cette façon une forme dégradée de phénoménologie, évoquant un monde de faits, de choses et d'états de choses. Nous avons donc maintenant deux extériorités, l'une idéelle, l'autre réelle (c'est ainsi que l'observateur s'exprime). Que peut encore la logique ? Offrir une explicitation tautologique en termes de conditions de vérité, c'est-à-dire paraphraser éternellement la norme intérieure dans un langage, à peine décalé, censé présenter l'extériorité du système comme déjà conquise.

Quelques variantes. Scinder par exemple la valeur en contenus étroits et contenus larges. Le contenu étroit est la valeur purement interne d'une forme, présumée descriptible dans un langage indifférent à la variété non pertinente des situations extérieures. Il s'agit d'une valeur conceptuelle générale, catégoriale. Le contenu large recouvre tout le reste, c'est une valeur située que le système rencontre sans pouvoir la déterminer catégorialement. Mais comme il n'y a pas ici l'équivalent d'une intuition et d'une imagination kantienne, mais seulement de la transduction, comment le système fera-t-il pour se soumettre cette valeur purement externe ? On ne voit même pas, d'ailleurs, comment stabiliser le contenu étroit dans l'indéfini des situations où il est appelé à valoir. Est-il donc encore un contenu ? Et si la systématique de ces contenus étroits reste logiquement assignable, ne devons-nous pas concilier à nouveau l'inconciliable, c'est-à-dire l'objectivité bien définie de la forme logique et l'indéfini de la nature englobante ? Non décidément, trop de logique (formes) et trop de conceptuel (valeurs) ne nous convient pas<sup>1</sup>. Et c'est ainsi que petit à petit espaces mentaux, images mentales, et prototypes divers sont venus enrichir la recette, et adoucir la transition de l'étroit au large. Mais pour cela, il fallait d'abord entrouvrir la porte à l'espace, mettre un peu de vague aux arêtes tranchantes de la forme logique, décloisonner les modules, bref changer de machine.

b) *Fonctionnalisme connexionniste*. — Se constitue sur le rejet de certaines valeurs sacrées du précédent : rejet de l'atomisme, du logicisme, de la compositionnalité (si tant est qu'il y ait de la constituance). Engagement en faveur du continu ; priorité à la microstructure de la cognition ; choix de mécanismes de type perceptif et associatif plutôt qu'inférentiel et logique. L'ontologie interne est libérée ; elle comprend maintenant au moins deux niveaux, dits micro- et macroscopique. Tout, y compris le temps, peut donc en principe être saisi sur une variété d'échelles<sup>2</sup>.

Reste qu'il s'agit toujours de représenter pour valoir quelque chose : le monde est alors un tissu de micro-traités, où se stabilisent des macro-configurations plus ou moins cohérentes. Autant dire qu'il ressemble comme un frère à ce réseau qui s'y adapte. Et cette ressemblance donnée dès le départ signifie que rien n'a changé vis-à-vis de l'extériorité, le

---

<sup>1</sup> Voir par exemple *Représentation et réalité* de H. Putnam.

<sup>2</sup> Pour plus de détails sur ce fonctionnalisme et sur le suivant, envisagés sous l'angle de la modélisation, je me permets de renvoyer à mon étude sur "Les modèles connexionnistes entre perception et sémantique" (*Sémiotiques*, Linguistique cognitive et modèles dynamiques, numéro spécial dirigé par J. Petitot, 1994, 6-7, p. 15-48

solipsisme étant toujours de rigueur. D'ailleurs l'ancien fonctionnalisme, quoique périmé, sera retrouvé comme cas-limite, grossière et utile caricature. Arrive donc le nouveau fonctionnalisme (il ne se présente pas d'emblée ainsi) qui va réformer et enrichir l'univers des formes – c'est-à-dire des représentations. Et puisque sa machine de référence est neuro-mimétique, de nouvelles relations plus fructueuses, plus équivoques, pourront se nouer avec les neurosciences.

Si donc la machine est un réseau connexionniste, le format de base est un espace vectoriel, l'espace de ses états, et toute forme s'individue comme un vecteur dans cet espace théoriquement de très grande dimension – en pratique beaucoup plus parcimonieux, hélas. L'observateur se préoccupe initialement d'une partie du réseau, les couches d'entrée et de sortie. Il cherche à comprendre comment telle ou telle association régulière entre les formes d'entrée et de sortie a pu s'établir, et sur quelles transformations intermédiaires elle repose. Pour cela, on ne définit pas de types formels, mais seulement des occurrences ; le réseau est en réalité une famille de processus, paramétrée par les poids des connexions ; on le règle par apprentissage (le terme de généralisation serait plus exact), à partir d'un échantillon de couples entrées-sorties. Si l'apprentissage réussit, il s'en déduit une leçon, moins empiriste qu'elle n'en a l'air au premier abord : telle architecture, tel codage d'entrée et de sortie, telle procédure d'apprentissage sont capables d'engendrer tel fonctionnement, en construisant sur les couches cachées des formes ou représentations intermédiaires caractéristiques de tout ce contexte, et ne prenant sens qu'à partir de lui. Permettre de penser cette contextualité (tâche, histoire, architecture) de l'efficacité des formes est certainement l'un des principaux apports de ce fonctionnalisme. Mais quel serait alors, du côté du monde, le répondant de cette contextualité inhérente, en quoi pourrait bien consister la perpétuelle relativité d'un tel monde ? Le positivisme implicite à cette problématique connexionniste a bien remplacé les “faits” de la construction logique par les “micro-traits” de l'environnement, mais en est-il plus avancé pour autant ? Un deuxième apport de ce fonctionnalisme doit toutefois être souligné : il a bien mis en évidence que le compromis réalisé entre contextualité et régularité tenait d'abord à la spatialité des représentations, ainsi qu'à la possible continuité des processus. Spatialité qui ne serait pas celle de l'espace physique, ni celle de l'espace d'une intuition sensible, mais celle d'un espace topologique abstrait dont les formes dépendent constitutivement. Il devient donc impossible de spécifier le rôle d'une représentation sans lui assigner une position, et les positions feraient alors système sur la base d'une topologie. Mieux encore, la possibilité de discriminer les formes et leurs rôles ne tient pas tant à la possibilité de les décomposer en



“parties”, comme dans le cas des formes logiques, mais d'abord à la possibilité de diviser leur espace entier en régions caractéristiques : la similitude des rôles formels ne renvoie donc plus à la possession d'un constituant commun, mais plutôt à l'appartenance à une même région de l'espace global.

Arrivent ensuite les objections, revers de cette contextualité dont nous parlions à l'instant. Ainsi Fodor et Pylyshyn : où sont donc passées la systématisme des processus et la compositionnalité des effets de sens ? Et cela marche : au lieu de répliquer en avançant résolument dans la direction d'une théorie gestaltiste et non compositionnelle de l'union entre la forme et le sens<sup>1</sup>, le nouveau fonctionnalisme confirme sa propre sujétion à l'ancien, qu'il prétendait dépasser. D'où un déluge de propositions visant à confirmer la capacité des réseaux à modéliser une constituance et une compositionnalité ressemblant à s'y méprendre à la précédente, avec en prime un peu de flou et de gradualité dans les réponses. Certes la primauté de la forme logique semble définitivement ébranlée, mais de façon contradictoire le noyau propositionnel de la grammaire tient bon, tout au moins sous la forme de représentations par traits des rôles casuels (agent, objet, patient, etc.). Ce noyau casuel-propositionnel résiste donc toujours aux tentatives de le compromettre avec l'espace et le mouvement, alors même que le néo-connexionnisme du groupe PDP s'était fixé pour tâche de comprendre ce qu'il y a de perceptif en toute opération cognitive.

Entre alors en scène le fonctionnalisme dynamique, qui pense non sans raisons pouvoir briser cette résistance.

c) *Fonctionnalisme dynamique*. — Bien plus ancien que le précédent, puisqu'il trouve son point de départ dans les oeuvres des mathématiciens R. Thom et Ch. Zeeman, il a dû néanmoins attendre le succès de son cadet connexionniste pour faire pleinement valoir ses titres sur la scène cognitive. J. Petitot est à mon sens, et de longue date, son premier acteur ; il en donne, année après année, les exposés les plus révélateurs, en le tirant toujours davantage, mais d'une façon créative, dans le sens d'un positivisme mathématique de la Forme que R. Thom désavouerait probablement<sup>2</sup>.

---

<sup>1</sup> Projet que l'on peut identifier, énoncé en termes assez peu voilés, dans un article de McClelland et Kawamoto (PDP, tome 2, p. 272-325, 1986).

<sup>2</sup> Voir notamment le livre *Mind as motion*, édité par Port et Van Gelder, qui contient un article-phare de J. Petitot (MIT Press, 1995). Cf. également *Physique du Sens* (1992), l'article “Phénoménologie naturalisée et morphodynamique” (*Intellectica* 17, 1993), et l'introduction au numéro 6-7 de la revue *Sémiotiques* (1994).

Ce fonctionnalisme se caractérise par son outil principal, voire constituant, qui est la théorie mathématique des systèmes dynamiques<sup>1</sup>, ainsi que par l'horizon actuel de son programme de recherche, à savoir l'ancrage du cognitif dans la perception visuelle. Il postule des fonctions symboliques, plutôt que des symboles internes, et fait de l'identification des rôles actantiels dans la situation en cours la première de ces fonctions. Elle consiste précisément en une structuration qualitative de l'espace perçu, à travers une perception catégorielle généralisée d'objets, de processus, d'événements, qui fonde une proto-symbolique ou morphodynamique cognitive très générale. Il s'agit en somme d'une sorte de visuo-cognitivism, qui cherche à reconstruire dynamiquement les diagrammes icôniques des linguistiques cognitives. Le noyau structural des grammaires n'est donc plus propositionnel, il est d'abord scénique. Les catégories discrètes, grammaticales ou conceptuelles, ne sont pas exclues, mais constituées à partir d'une co-articulation constante au monde perçu, à travers une schématisation topologique des interactions spatio-temporelles.

En tant que philosophie naturelle de la forme, ce fonctionnalisme l'emporte facilement sur les précédents. Il offre en effet des contreparties objectives (mathématiques) à l'idée de synthèse formelle, à la détermination réciproque du global et du local. Il offre, pour ainsi dire gratuitement, une modulation indéfinie des formes, en même temps qu'une méthode générale pour réduire leur complexité, pour ensuite les catégoriser. Il les construit par esquisses, leur confère un temps intrinsèque d'intégration et de stabilisation qui les inscrit dans un Présent doté d'une épaisseur, et non dans le temps extrinsèque d'une horloge. En tant que dynamiques, ces formes sont d'emblée en connivence avec une physique des substrats neuraux, d'une part, avec une phénoménologie de la vision, d'autre part. Elles ne sont pas soumises à une théorie de l'instanciation à l'identique des types en tokens, car elles se déduisent d'un potentiel (d'un champ de vecteurs, par exemple), différent par construction de chacune des actualisations qu'il détermine. Les systèmes du fonctionnalisme dynamique pourraient aussi se faire adaptables, et dans ce cas les potentiels de formes évolueraient eux-mêmes à la faveur de leurs actualisations. Il arrive même qu'on les implante sur des réseaux connexionnistes, ce qui leur donne, pense-t-on, un surcroît d'effectivité.

Quelle est alors la machine de ce fonctionnalisme, et en quoi d'ailleurs est-ce toujours un fonctionnalisme ? On sera tenté de répondre qu'il n'a pas de machine qui le caractérise, car tout espace et tout système

---

<sup>1</sup> À laquelle il convient d'ajouter (au moins) la théorie des singularités des fonctions différentiables, ainsi qu'un corps récent de mathématiques du traitement d'image.

dynamique sont *a priori* une de ses machines, définissent un de ses formats. Et c'est précisément l'équivoque toujours maintenue entre l'espace du substrat physique et l'espace des formats, entre l'espace physique englobant et l'espace perçu, qui détermine cette philosophie comme un fonctionnalisme. L'effectivité des processus concerne une perception toujours solipsiste, et les rétines virtuelles qui tiennent lieu ici de surfaces photo-sensibles sont en réalité des espaces déjà phénoménologiquement valués, ce ne sont pas des espaces physiques authentiques, chevillés à un organisme mobile. Ils supposent déjà réussie une fois pour toutes une certaine stabilisation spatiale, sans soumettre en permanence sa reconstitution à l'épreuve du mouvement réel. Les formes sont donc construites sans référence à l'action qui signe leur valeur, et leur espace n'est pas constitué de mouvement, intrinsèquement valué de mouvement. De ce défaut dans le couplage constituant à la motricité et à l'extériorité résultent nécessairement deux conséquences. L'une, qui s'énonce plus adéquatement en termes kantien, serait qu'on a ainsi confondu l'imagination (en tant que schématisation dans une intuition pure) avec l'intuition empirique proprement dite, ou alors qu'on a totalement subordonné la seconde aux *a priori* de la première, conçue à la façon d'un cadre existant à l'état séparé. L'autre conséquence serait qu'on manque l'idée (husserlienne) de perception comme acte du vivant dans son monde, et l'idée de forme perçue comme corrélat de cet acte. On perd toute chance de comprendre comment l'action initiée détermine cela-même qui est à percevoir, tandis que la perception toujours recommençante suscite et focalise l'action ; comment, par conséquent, une valeur est déjà reconnue et active avant même que la forme ait fini d'advenir. Si bien qu'on en revient nécessairement à l'idée typiquement fonctionnaliste d'une intentionnalité fondée sur la référence à un monde donné par avance, achevé avant qu'on se tourne vers lui. Que les Gestalten soient mathématisées n'y change rien, elles restent des représentations, rapportées ici à une théorie du reflet ou *mimesis* entre la forme de l'intériorité et celle de l'extériorité (une théorie qu'on pourrait qualifier de pré-cartésienne).

Le fonctionnalisme dynamique est donc bien un fonctionnalisme. Sa "machine" est un espace cognitif purement intérieur, considéré dans sa plus grande généralité. C'est *l'espace fonctionnaliste*, en tant que forme mathématique universelle de la possibilité de tous les formats, considérés indépendamment de leur turbulente matière. C'est aussi *l'espace fonctionnaliste*, en tant que cadre invariant de l'inscription des formes, donné préalablement à leur variation au lieu d'être construit par la variation elle-même. C'est encore *l'espace fonctionnaliste*, en tant que scène

intérieure de la représentation d'actions, qualifiées par un “système ” pourtant dépourvu des moyens et du désir de s'y impliquer.<sup>1</sup>

Mais il y a un autre aspect de la généalogie du fonctionnalisme dynamique qui fait transition vers les fonctionnalismes suivants, car il pointe sur la biologie, plutôt que sur la psychologie. Il s'agit de la notion, centrale dans la sémiophysique de R. Thom, de “prégnance ”, en tant que valeur pulsionnelle projetée, “ propagée ” par l'organisme sur une forme ou “saillance ” constituée au préalable. La prégnance elle-même n'est pas une forme, mais un potentiel d'accrochage aux formes qui les rend signifiantes, au sens où il commande une attitude spécifique (facilitation sensorielle, préparation motrice) chez l'animal. Même si le traitement qu'en donne R. Thom s'avère plutôt décevant, il paraît difficile de ne pas y voir en effet une modalité fondamentale de la relation à l'*Umwelt*. De son côté, le fonctionnalisme dynamique, qui centre pour le moment son effort sur un mode “proto-symbolique ” et peut-être spécifiquement humain de la cognition, s'est jusqu'ici abstenu de faire appel à quelque valeur analogue, émotionnelle ou autre ; si bien que ses reconstructions de scènes nous semblent concerner de simples mouvements, et non des actions à proprement parler, dont la reconnaissance exigerait probablement une forme ou une autre d'empathie. Toujours est-il que j'évoque ici la notion de

---

<sup>1</sup> On dira peut-être que tout cela n'est qu'un début, et qu'il faut bien simplifier si l'on veut commencer. On rappellera aussi Galilée et son plan incliné. Peut-être... Toutes les simplifications sont les bienvenues en effet, si l'on reste suffisamment lucide, scientifiquement et philosophiquement, sur ce qu'elles nous font connaître. Thom et Chomsky, si différents soient-ils, ont au moins quelque chose en commun : un art, discutable et génial, de la simplification des systèmes complexes. Ils ont tous les deux proposé des réélaborations de la distinction entre *description* et *explication*, en intégrant, chacun à sa manière, des déterminations “dynamiques ” ou “génératives ” formelles à leurs dispositifs théoriques. Et ils ont tous deux maintenu une distance protectrice entre leurs travaux et les interprétations psycholinguistiques ou biophysiques effectives qu'ils ont pu susciter : Thom en se désintéressant généralement de ces interprétations, et Chomsky en les soumettant aux idéalizations préalables de la *compétence*. L'informatique ne fait que confirmer ce tableau comparatif, en construisant des processus qui spécifient davantage les idéalités de la morphosyntaxe, tout comme celles de la morphogenèse qualitative à la Thom. Mais ne prenons pas ces processus pour des *bases* nécessaires ou suffisantes de nos propres performances (à quelque niveau qu'on prétende les ressaisir). Célébrons ainsi tant qu'on voudra les théoriciens et les ingénieurs qui construisent des machines à segmenter-et-structurer des phrases, ou plus récemment des images. Et comprenons du même coup l'analogie entre les processus morphosyntaxiques (de la linguistique computationnelle) et les processus morphodynamiques (de l'imagerie computationnelle) : ils présentent, dans leurs domaines respectifs du “langage ” ou de la “vision”, le même intérêt, la même portée, et les mêmes risques de mésinterprétation.

prégnance parce qu'elle illustre encore une fois l'idée que les formes seraient constituées avant les valeurs qui s'y attachent, que les saillances se manifestent donc avant même de focaliser l'attention, pour ainsi dire dans une sorte de vide intentionnel, hors de toute orientation de l'organisme<sup>1</sup>. C'est justement cette conception des formes en tant que données préalables (c'est-à-dire préalables à leur évaluation) que l'on retrouve dans le cadre des biofonctionnalismes.

d) *Biofonctionnalismes*. — Alors que les trois fonctionnalismes précédents étaient nommés à partir de leur instrument d'objectivation privilégié, le terme de biofonctionnalisme renvoie au vivant lui-même, en tant qu'il dépendrait pour sa survie d'un certain concours de fonctions. L'une d'entre elles est évidemment la capacité à représenter, probablement restreinte à certaines structures matérielles du système nerveux<sup>2</sup>. Mes connaissances sur ces biofonctionnalismes étant des plus restreintes (d'autant plus qu'ils ne sont pas, à ma connaissance, solidaires de courants de modélisation bien précis<sup>3</sup>), je me bornerai à pointer ce qui en fait encore, selon mes critères, des fonctionnalismes.

C'est qu'en effet on pourrait penser avoir fait avec eux un pas décisif vers la sortie. On reconnaît d'abord une valeur suprême, survivre en tant qu'espèce, qui interdit la moindre suspension de l'extériorité. Les fonctions en sont dérivées, selon un principe dont la version apparemment la plus perfectionnée stipule qu'on les attribue à toute structure matérielle

---

<sup>1</sup> Il est possible cependant que pour R. Thom, la prégnance ne soit qu'une partie de la valeur intentionnelle globale projetée par l'animal vers la saillance. Mais cela laisse intact le problème de l'intentionnalité responsable de la saillance préalable. Peut-être pourrait-on invoquer ici une sorte "d'archi-prégnance" non différenciée qui investirait toute saillance comme cible intentionnelle, avant ou en même temps que d'autres prégnances plus différenciées.

<sup>2</sup> Alors que pour R. Thom, par exemple, l'endoderme de l'embryon représente aussi ses proies futures. Rappelons que selon le biostructuralisme thomien, il y a isomorphisme (en un sens très abstrait) entre la structure organique, tissulaire de l'animal, et ce que Thom appelle sa *figure de régulation* (c'est à dire sa forme de vie), qui est une figure dynamique déployée dans un espace – immense mais diversement factorisable – de description des conduites et des schèmes sensori-moteurs (comme respirer, se nourrir, chasser, etc.). Je suppose que dans ces conditions, le cerveau ne représente rien d'autre que la coordination des champs tissulaires différenciés, ou isomorphiquement celle des conduites.

<sup>3</sup> Les courants du type *Vie Artificielle* ou *Animats* occupent peut-être le même terrain imaginaire, mais n'entretiennent à part cela aucune relation organique avec les biofonctionnalismes. Ils ont d'ailleurs produit à plusieurs reprises des manifestes anti-représentationnalistes (ceux de R. Brooks, par exemple). Reste que ces courants scientifiques sont eux-mêmes partagés, et souvent peu clairs sur leurs présupposés.

reproductible responsable (toutes choses étant égales par ailleurs) d'un effet conférant un "avantage" à l'espèce dans la vaste économie darwinienne de la Nature (où toute valeur provient, non de la force de travail, mais de celle de reproduction). Les fonctions ne sont donc plus *identifiées* à partir de différences de rôles purement *internes* aux systèmes, mais d'abord à partir des différences constatées *d'un individu à l'autre*, en tant qu'assignés à la même obligation de se reproduire. On ne saurait surestimer l'importance de ce décalage dans la définition, et l'on peut même penser que les fonctions des téléofonctionnalismes sont de simples homonymes de celles qui les ont précédées .

En tout cas, nulle trace apparente de machine dans ce dispositif. Une dépendance instituée dès le départ entre les fonctions et l'entour du système. Où sont donc les séparations et les priorités caractéristiques des fonctionnalismes ? Cependant :

– S'il n'y a pas explicitement de machine intérieure aux individus, il semblerait que la Nature toute entière se soit mise à calculer, car sa transformation est maintenant indexée sur une distribution "d'avantages" inter- ou intraspécifiques. Mais à supposer que l'on ait bien déterminé ces avantages, comment fera-t-on pour déterminer ensuite les fonctions recherchées ? comment contourner la "causalité circulaire" propre à toute clôture systémique, c'est-à-dire la dépendance réciproque des causes et des effets, qui fait de tout une cause et un effet à la fois ? comment localisera-t-on la structure productrice d'un "effet", si cet effet ne se définit pas simplement en termes de bilan métabolique ou mécanique local ?

Plus encore, il est bien difficile de saisir ce qu'est l'avantage évoqué, et comment l'estimer. Si le prédateur augmente en appétit et en puissance, qu'il épuise ses ressources en proie, et disparaît à son tour, où est l'avantage ? Mort, où est ta victoire ? Qu'est ce qu'une "optimisation", si elle se transforme en catastrophe ? La question de la portée temporelle de ces avantages est donc décisive. Mais on s'étonne alors de la relativité de fonctions qui en dépendraient à ce point, qui seraient à ce point relationnelles : qu'aurait-t-on alors gagné à tenter une définition, qui n'est pas seulement inspirée par un souci heuristique tout aussi relatif, mais faite à l'évidence avec l'intention de localiser et stabiliser dans une intériorité la complexité d'une relation changeante à l'extériorité ? Et quand bien même on aurait, par quelque méthode que ce soit, isolé un noyau fonctionnel dans l'organisme et mis le reste en chômage technique, est-on vraiment sûr que la survie ne dépende pas de ce volant de réserve ? Ne vaudrait-il pas mieux, en somme, pour être un meilleur fonctionnaliste, partir d'une théorie des "formes de vie" ressaisies du point de vue de leur

autonomie, pour en dériver ensuite les descriptions fonctionnelles ? Ce serait là engager une phénoménologie du vivant toute différente, car une forme de vie comporte par essence l'*auto*-reproduction, c'est la forme de l'être-au-monde d'un individu *singulier*, et non pas seulement une forme générique de l'espèce en tant que globalement déterminée par le survivre-et-se-reproduire de ses membres. On n'en retrouverait pas moins les fonctions de la physiologie classique (respirer, se nourrir, etc.) en tant que distribuées sur un collectif organisé de tissus. Une modélisation de la morphologie et du comportement pourrait alors suivre, sans pour autant être gagée sur une *ontologie* des fonctions, et sans oublier ce qu'elle devrait à cette description préalable des formes de vie comme acteurs de l'évolution .

– Il y a, ou plutôt il y a eu, des fonctionnalismes qui prétendaient apporter une ontologie du *mind*, de l'esprit, mais non une épistémologie, c'est-à-dire une justification de la façon de connaître et déterminer empiriquement cette ontologie (pensons par exemple aux inconnaissables processus centraux dans la *Modularité de l'esprit*). On pouvait même risquer un pronostic : ou bien un fonctionnalisme partirait des valeurs pour aller aux fonctions, et devant la difficulté d'aller plus loin, produirait simplement, sur des bases finalistes, une ontologie des valeurs et des fonctions sans épistémologie ; ou bien il partirait des formes pour aller aux fonctions, prétendrait en construire aussi l'épistémologie, mais s'en tiendrait là, demeurant dans le formalisme et le solipsisme, et renonçant à ontologiser le passage à la valeur. Mais les biofonctionnalismes en veulent bien davantage : il ne s'agit pas seulement de cerner ontologiquement les constituants des noyaux fonctionnels, mais plus encore de déterminer empiriquement les fonctions. N'est-il pas question, par exemple, de “resserrer les liens entre fonction et causalité, et donc en particulier, de déterminer les contraintes physicalistes qui s'attachent à l'accomplissement de la fonction représentationnelle” ? n'est-il pas question aussi de quantifier des valeurs adaptatives, et de faire jouer une optimisation sur l'espace des phénotypes possibles ? S'il s'agissait simplement de modélisation, prise comme heuristique ou expérience de pensée, je n'y verrais aucune objection. Mais l'objectif scientifique s'efface ici devant l'entreprise ontologique, qui cherche à donner des bases substantielles déterminantes à de simples principes de réflexion. Estimer des valeurs adaptatives sous forme d'indices numériques aux fins de modéliser est une chose ; donner à ces valeurs un statut ontologique en est une autre, qui relève d'un positivisme peu justifiable<sup>1</sup>.

---

<sup>1</sup> On retrouve la même démarche positiviste dans l'utilisation du concept de probabilité par Dretske : alors que c'est un concept relatif à un *ensemble* d'événements en

– C'est enfin sur la question des formes, ou phénotypes, que se manifeste de la façon la plus flagrante le caractère authentiquement fonctionnaliste de ces philosophies. Les formes ne sont-elles pas données hors de toute épigénèse, et offertes toutes montées à l'impitoyable Sélection ? On voit même, parfois, l'insatiable ogresse extraire ses victimes, toutes ses victimes potentielles, d'un répertoire... *fini*, où le misérable cheptel était rassemblé. De ce que l'ensemble des génotypes est fini, on déduit la finitude de celui des phénotypes : mais s'agit-il des phénotypes virtuels, ou de ceux que l'ontogénèse construit effectivement ? Séparer la question de l'ontogénèse de celle de l'organisation empêche de concevoir l'organisation comme possibilité de la réorganisation permanente, et donc comme reproduction permanente de cette possibilité. Une telle séparation ouvre la voie à une réduction ou mise au format de la diversité biologique, en laquelle on reconnaîtra l'analogue des mises au format des fonctionnalismes "machiniques" traditionnels. Elle suppose aussi un accès aux formes indépendant de la matérialité de leurs substrats. Elle suppose enfin une ontogénèse des formes indépendante des valeurs, ou prégnances, constitutives de l'habitus du "système" vivant, une ontogénèse indépendante du "fonctionnement" pris dans sa matérialité. Les fonctions, en tant que rapports de production stabilisés, l'emportent alors sur la morphogénèse : et l'organe, ou le tissu, n'est plus que la prothèse au service de la fonction qui l'enrôle, le simple substrat de la fonction qu'il incarne au présent en tant que propension<sup>1</sup>.

Les biofonctionnalismes, en somme, conjugueraient une métaphysique fonctionnaliste à une erreur aussi courante que profonde sur le sens des explications darwiniennes. Partant du fait qu'il y a de la Vie et de la Forme, une explication darwinienne nous indiquera seulement pourquoi telle vie et telle forme effectivement rencontrées étaient possibles, ou plutôt pourquoi elles n'étaient pas aussi impossibles que certaines variantes imaginaires ou réelles (qui ont dû céder le pas). Cette explication établira donc l'impossibilité de certaines variantes, ou *a contrario* les bons atouts de certaines autres. Présupposant la vie et les formes, elle en fera

---

concurrence, qui doit être déterminé au préalable, Dretske n'hésite pas à parler de probabilité d'un fait, comme si le monde était *l'ensemble* (déjà déterminé) de tous les faits.

<sup>1</sup> Voir sur ces questions le numéro 16 d'*Intellectica* coordonné par J. Stewart sur *Biologie et cognition*. La parenté entre les fonctionnalismes et les néo-darwinismes y est analysée tout du long, dans une perspective critique que Stewart nomme pour sa part une perspective *constructiviste*. L'article de Charles Lenay, notamment, analyse très précisément ce qu'il en est de la parenté et des alliances de ces paradigmes biologiques ou psychologiques à partir du sort commun qu'ils ménagent aux notions de *forme* et de *représentation*.



comprendre les variations, mais ne les engendrera pas. Jamais elle ne nous expliquera causalement, d'une façon déterminante, qu'il puisse y avoir ne serait-ce qu'un seul être vivant ou une seule forme. Un raisonnement darwinien ne détermine donc pas effectivement la vie, mais éclaire plutôt la plausibilité de la vie effective en repérant ses variantes impossibles ou éliminées. En aucun cas il n'explique l'existence du "générateur de diversité" que présupposent ses reconstitutions (invoquer ici l'ADN ne sert à rien). Lorsque par conséquent on utilise ce type d'explication, on présuppose toujours, quoique de façon parfois dissimulée, l'existence de l'être dont on veut estimer la possibilité d'avoir été. On ne peut donc s'en prévaloir pour déterminer causalement cet être, le mener à l'existence au sein d'un schème causal mécaniste<sup>1</sup>. En termes kantien, les explications darwiniennes s'appuient toujours sur des jugements *réfléchissants*, puisqu'elles reconstituent les phénomènes à partir de certains principes de réflexion, comme celui qui analyse la vie à partir d'une "obligation" de se reproduire ; en aucun cas ces explications ne peuvent se réduire à des jugements simplement *déterminants*, du type de ceux qui assignent les phénomènes à l'existence à partir d'un schème strictement causal. De la même façon, nous ne pouvons confondre deux projets, même s'ils sont tous deux parfaitement louables : d'une part celui de *spécifier* les systèmes solidairement de leur entour (un pas en avant important à porter au crédit des biofonctionnalismes), et d'autre part celui, bien différent, de *déterminer* matériellement, dans l'espace et dans le temps, leur *genèse* au sein de ce même entour. La perspective darwinienne sur l'évolution introduit bien du *relationnel* dans la nature, en tant qu'elle est considérée, selon les propres termes de Darwin, comme une "économie". Et ce que les biofonctionnalismes, en tant que philosophies de la naturalisation, essaient de faire, c'est précisément de *déterminer* l'espace et le temps de la nature à partir de cette relation, dont ils font un principe *philosophique* de factorisation de la complexité physique et biologique<sup>2</sup>. Or la physique

---

<sup>1</sup> Il existe bien des modèles darwiniens internes de l'épigenèse, qui soumettent au même type de raisonnement des populations de cellules, de synapses, d'assemblées, etc. (cf. le darwinisme neuronal de Changeux ou Edelman). Ils s'inscrivent certes dans une causalité physique, mais toujours en se donnant au départ sans l'expliquer la diversité et la cohérence biologique au sein de laquelle on reconstruit la réorganisation des systèmes. C'est à ce prix qu'on montre comment un organisme, du moment qu'il ne meurt pas et conserve en toile de fond son unité et sa diversité intérieure propres, peut "effectivement" passer d'un état A à un état B.

<sup>2</sup> Il s'agit évidemment d'un principe de compréhension extraordinaire, mais là n'est pas la question. Car c'est de philosophie que nous parlons, et nous n'avons pas besoin que la philosophie construise des ontologies (nous avons déjà en sciences toute

se constitue précisément en interdisant la possibilité d'une telle détermination aberrante (de l'espace par quelque chose qui ne serait pas d'emblée spatial, mais d'abord une relation). Si bien que la nature des fonctionnalismes ne peut être une nature physique, mais seulement une nature métaphysique ou “ontologique” : soit une simple *image* de la nature, engendrée par l'usage illégitime (constituant, dirait Kant) de certains principes de réflexion – fort précieux par ailleurs. Si bien que la clarification – souvent remarquable – de ces principes est immédiatement obscurcie par l'appel à la causalité, au “physicalisme”, et peut-être par le désir de faire de la science sans en faire, sur un mode relationnel, précisément.

*Une remarque sur Thom et Darwin.* — A l'image (évoquée ci-dessus) d'une économie de la Nature, un raisonnement néo-darwinien pourra préférer celle d'une résolution de problèmes. L'heuristique qui court alors en filigrane consiste à éclairer les situations en les soumettant à la tension d'un problème qui s'y poserait (ou se serait posé dans un état antérieur), pour retrouver l'existant comme solution à ce problème, inventée par les coups de dés de l'évolution. Il est très instructif de comparer cette démarche à celle de R. Thom, qui fait fond, quant à elle, sur une détermination a priori des formes, relativement à l'espace et au temps où leur stabilité doit se manifester. On pourrait résumer sa méthode de la façon suivante : “Tout ce qui fait sens pour nous, pour autant que nous puissions le connaître, doit pouvoir se déterminer dans une actualisation spatio-temporelle (réduction du sens au phénomène). Or l'apparaître du phénomène se résout pour l'essentiel dans une donation de formes stables, ou plus dynamiquement dans des interactions de formes à l'instabilité drastiquement limitée (réduction du phénomène à la forme). De plus les concepts de stabilité, ou d'instabilité d'ordre fini, se laissent construire mathématiquement (théorie des catastrophes généralisées). Il s'en déduit un principe d'intelligibilité des phénomènes, qui consiste à chercher d'emblée à les décrire, puisqu'ils existent et nous apparaissent, dans les termes d'une schématisation mathématique qui ne “sélectionne” par construction que des formes suffisamment stables”. Naturellement, une telle approche n'expliquera pas pourquoi il y a de la stabilité plutôt qu'autre chose. Mais partant du principe qu'il y a de la Forme, et que l'Etre ne peut se manifester autrement qu'en se stabilisant comme forme, elle nous en fera comprendre certaines variations. Exprimée de cette façon, l'herméneutique thomienne se révèle beaucoup plus compatible à l'herméneutique néo-darwinienne qu'on ne pourrait le croire, si l'on s'en rapporte aux violentes controverses qui les ont opposées. On peut d'ailleurs combiner leurs principes de réflexion spécifiques : il suffit par exemple de réfléchir à la situation en imaginant que le “problème” posé à l'évolution est justement la génération d'une forme stable (c'est-à-dire résistante aux perturbations présentes de la situation), et en même temps réitérable (c'est-à-dire reconstructible à travers une gamme indéfinie de situations). E. Bonabeau et G. Theraulaz ont ainsi proposé des modèles de construction des nids d'insectes sociaux qui relèvent, me semble-t-il, d'une problématique toute proche : ils

---

l'imagination nécessaire), et pas davantage besoin qu'elle nous propose des simplifications (car nous ne lésinons déjà pas là dessus).

définissent d'abord un schème de comportement individuel comme un ensemble bien déterminé de couples stimulus-réponse, qui ne met en jeu que des configurations locales ; lorsque tous les agents appliquent le même schème de comportement bâtisseur, on obtient des comportements collectifs de type *stigmergique*, où l'état de la construction en cours, appréhendé localement par chaque agent, suffit à orienter correctement la poursuite globale de l'activité ; les “nids” sont, parmi toutes les formes obtenues de cette façon, celles dont la morphologie reste qualitativement semblable (stabilité à la Thom) lorsqu'on réitère la construction en appliquant toujours le même schème (les résultats peuvent différer, car les déplacements des agents sont en partie aléatoires) ; les auteurs recherchent alors les seuls schèmes de comportement admissibles, à savoir ceux qui engendrent de façon robuste, à chaque réitération, une forme qualitativement invariante (donc un nid, par définition). Il y a bien ici croisement – ou presque – entre Thom et Darwin : d'abord parce qu'il s'agit d'un modèle qualitatif et dynamique de morphogenèse ; et ensuite parce que les composantes stochastiques du modèle (les mouvements aléatoires des agents) pourraient s'interpréter indifféremment et simultanément de plusieurs façons : comme fluctuations instantanées de l'environnement, comme diversité des ontogenèses individuelles, comme diversité des génotypes, etc. Il importe peu, finalement, que ces variations aléatoires soient ou non des variations génétiques, qu'elles s'exercent plutôt en synchronie (à l'intérieur de l'essaim, ou bien d'un essaim à l'autre de la même espèce) ou dans la dérive diachronique de l'espèce : pour ce type de modélisation qualitative, en effet, on peut rassembler “synchroniquement” toutes ces dimensions de variation dans un seul et même espace global de paramètres, dont chaque vecteur détermine une variante formelle d'un processus de “morphogenèse” généralisé (engendrant un organisme, un nid, un comportement individuel ou collectif, etc.). Le modélisateur peut ainsi déterminer à sa convenance la vitesse, l'amplitude, et les aleas des perturbations paramétriques auxquelles la forme étudiée doit qualitativement résister pour être dite stable (il n'en irait pas de même si l'on souhaitait construire une image plus effective des processus, car il faudrait alors régler les constantes de temps de façon moins arbitraire). Encore une fois, un tel modélisateur darwino-thomien n'ira pas croire qu'il fait de la physique, par exemple, sous le prétexte qu'il en reprend les concepts et les méthodes. Il se considèrera plutôt comme un enquêteur, qui part d'un constat (il y a des formes stables et réitérables qui caractérisent les espèces), pour en inférer des déterminations qualitatives sur la gamme des formes et des morphogenèses possibles (et corrélativement sur les perturbations paramétriques compatibles avec une forme donnée). Voir les articles de Bonabeau et Theraulaz dans *J. of Theor. Biol.* et dans *Science* (1995), et le livre qu'ils ont coordonné sur *L'Intelligence collective* (Hermès, 1994).

5. *Postulats, exercices.* — Nous avons donc isolé le noeud de notre différend avec les fonctionnalismes, que l'on entende sous ce nom une variété académique de philosophie ou bien une sorte d'épistémologie, plus ou moins spontanée, interne aux sciences cognitives.

En premier lieu ces philosophies – ou ces épistémologies – permettent toujours, et même recommandent, que les formes soient constituées, dans leur formalité de référence, sans dépendre constamment et avant même

leur complète advenue, de principes d'évaluation eux aussi en voie de formation ; et sans dépendre, si ce n'est de façon contingente, d'actions opérées dans l'entour du système. En second lieu, elles tentent de réparer cet irréparable défaut initial en stabilisant et ontologisant ces différents modes de la description, soit en construisant d'abord une ontologie-pivot de fonctions, soit en déterminant une ontologie de référents indépendants de tout être-au-monde.

Nous défendons ici des postulats opposés. Valeurs, formes, actions, fonctions ne peuvent que se co-constituer, et aucune ne peut commencer à part des autres. Elles doivent donc advenir dans le temps, leur temps doit être intrinsèque à leur définition. La première dimension à "internaliser", si tant est que cette phrase ait un sens, c'est le temps. Notre intelligibilité se construit alors en insérant des décalages temporels "infinitésimaux" dans cette détermination réciproque, et en l'examinant successivement de différents points de vue dont nous ne pouvons comprendre, ou plutôt imaginer, l'intégration sinon par l'analogie offerte par la modélisation. Et si nous ne pouvons encore comprendre de cette façon, en conjuguant le temps à la détermination réciproque, nous devons en tout cas toujours *décrire* les phénomènes. En cherchant à réduire, naturaliser, normaliser, sans inscrire toutes ces tentatives dans une herméneutique, on ne rend aucun service à la science. On ne lui suggère par là aucun objet, on la tire en arrière à la recherche d'ontologies ; tandis qu'on fait de la philosophie un simple rêve de laboratoire, ou bien une idéologie, c'est-à-dire une justification des intérêts et des tractations symboliques jouant dans la pratique scientifique.

C'est à peine, d'ailleurs, si nous pouvons comprendre encore la démarche sidérante des sciences cognitives : elles ont constitué comme séparé leur propre réseau d'identités disciplinaires (ce qui est parfaitement légitime), mais en prétendant, ce faisant, avoir défini la mesure principale de ce qu'est l'humain, ou même le vivant en tant que porteur de conduites, objet possible d'une psychologie. Le fonctionnalisme est alors l'espèce philosophique qui s'est développée à l'intérieur de cet étrange écosystème, dans l'oubli de toute anthropologie, et bien longtemps de l'éthologie. On a ainsi gagé toute valeur sur la capacité de référer, mais à quel monde ? Comment espérer dans ces conditions qu'un accord, ou du moins un réseau de passages, s'établisse entre la valeur (prise comme affect, émotion, sentiment de plaisir ou de peine), la valeur encore (comme connaissance réussie), la valeur toujours (comme qualité pratique des intentions dans l'action), ou encore la valeur (comme constitution au miroir d'autrui et différence d'avec lui) ? Valeur n'est pas le nom d'un étant, mais celui d'un renvoi plus ou moins réussi, plus ou moins harmonieux, entre ces différentes capacités d'agir, de

sentir, etc., se formant les unes les autres dans leur exercice effectif, plutôt qu'elles ne s'informent sur la base de distinctions définitives.

Ainsi l'information, par exemple, ne peut être qu'une idéalisation de ce à quoi un processus de réorganisation permanent couple l'organisme ou le système ; l'information ne préexiste pas à des systèmes qui la traiteraient en s'y conformant, elle émerge seulement, relativement à un observateur, comme distinction fonctionnelle ou environnementale pertinente dans ce couplage. Si on y fait appel, ce ne doit pas être sur le mode ontologique qui la fait s'instancier dans la matière vivante, après avoir préexisté séparément dans une intuition pure, mais comme un concept ou schème qui ne se substitue nullement aux phénomènes dont il nous aide à énoncer les lois.

Ainsi également des intentions d'agir : dira-t-on que ce sont elles qui, de la forteresse intérieure où elle résident, lancent une action planifiée au préalable à la poursuite d'un objet bien calibré et déjà constitué dans l'extériorité ? Nous dirons plutôt que l'action-perception est le déploiement interprétatif de l'intention initiale, qui n'est encore qu'une amorce ou esquisse partie à la recherche de sa propre définition dans le cours d'action ; que l'intention est donc la réserve de son supplément essentiel d'action, où elle se découvre comme possibilité d'insister, de se rectifier ; qu'elle est aussi la faculté de converger, de trouver ou déposer sa signification dans l'objet qu'elle rencontre comme sa réussite provisoire ; qu'elle est aussi la capacité de revendiquer son accomplissement et de se relancer, etc<sup>1</sup>. Intention, action-perception, et objet naissent, vivent, meurent et ressuscitent ensemble, ce sont des moments d'un même procès intérieur et extérieur à la fois, semblable au co-déploiement du signe, de l'interprétant et de l'objet dans la sémiotique peircéenne – du moins quand ce ne sont pas Morris et Carnap qui la reformulent<sup>2</sup>. Quelque chose d'assez différent, donc, d'un signe intérieur renvoyant à un référent extérieur par la médiation d'un concept intérieur, différent aussi d'une attitude propositionnelle possédant un certain contenu pointé sur un référent.

Mais il y a plus encore que ce postulat selon lequel l'intentionnalité ne peut être détachée de la matérialité du couplage entre une intériorité et une extériorité en détermination réciproque. Car nous pouvons, selon les besoins d'une négociation toujours à reprendre, moduler la part concédée

---

<sup>1</sup> Je renvoie ici aux belles analyses de P. Livet dans sa *Communauté virtuelle* (1994, Editions de l'Eclat, notamment le chapitre 2, sur l'action en cours de définition).

<sup>2</sup> Mais plutôt C. Tiercelin, dans *C. S. Peirce et le pragmatisme* (P.U.F., 1993), et *La Pensée-signe* (J. Chambon, 1993).

à l'intérieur comme à l'extérieur. Nous pouvons par exemple centrer notre attention sur les configurations extérieures, s'il s'agit d'une activité finalisée par une tâche pratique, s'exerçant dans un espace dédié : bureau, cuisine, atelier... Nous y verrons un cas d'action et de cognition situées, et ferons des objets le support d'*affordances*, c'est-à-dire de suggestions d'agir. Ou mieux encore, nous penserons à Peirce plutôt qu'à Gibson, et dirons que la situation est *sémiotiquement* perçue, puisque l'action est son interprétation. Le sujet nous apparaîtra alors comme une certaine puissance de commutation ou de passage d'une configuration matérielle à une autre. Chaque configuration, en tant que sémiotique, vaut pour mémoire externe de la tâche, trace active de ses acquis et de son projet, qu'il faut sans cesse réélaborer et réécrire<sup>1</sup>. Ce faisant nous n'avons pas annulé le psychisme, la réserve intérieure, nous avons au contraire étendu l'esprit à la matérialité des configurations d'objets dont nous dirons sans rire que la saillance s'accroît en même temps que les dispositions du sujet à agir.

Ainsi encore des analyses finalistes ou fonctionnelles. Je suis par exemple assis à mon bureau, en train d'écrire : une table, un ordinateur, une lampe. Est-ce là toute la scène de l'écriture ? Pas tout à fait. Pour avoir la scène, il faut que je recrute (mais en réalité c'est déjà fait) les objets, les saillances, au service de la figure globale et finale : l'ordinateur est pour-écrire, la table est aussi pour-soutenir la-lampe, qui éclaire le tout. J'ai donc constitué un Présent sous la tension d'une certaine fin, distribué des rôles dans une interaction instrumentale, inscrit une activité qui vaut au-delà de ce Présent. La scène est donc déjà un signe, ce qui est bien dire que sa valeur n'est pas seulement une modalité de l'extériorité donnée en présence, mais d'abord une disposition (fragile, incertaine) de ma relation à elle. Je peux alors prendre cette scène comme point de référence, et me promener dans le temps. Au lieu de voir le présent comme le fils soumis du passé, je recrute le passé au service du présent, et me demande ce qu'il fallait pour en arriver là, ce qui aurait pu empêcher qu'on y vienne, en quoi cette scène est véritablement une rescapée, d'entre mille autres virtuelles et définitivement naufragées. En réfléchissant *suivant ce principe* au présent que j'affronte, je reconnais aussi bien la persistance des actes et des intentions, que la simple conformité matérielle des objets aux fins sous lesquels ils sont enrôlés (temporairement, par moi). Mon attention est donc attirée sur certaines déterminations matérielles, dont je comprend qu'elles concourent à la dynamique du processus sans pour autant le déterminer : force de

---

<sup>1</sup> Cf. les travaux de B. Conein, D. Kirsch, D. Norman, Hutchins, L. Suchman, J. Theureau...

l'ampoule, hauteur du siège, tablette de chocolat, trop de Bordeaux. Grâce à ce principe de réflexion, que je peux aussi bien appliquer en direction du futur, j'affine ma perception de ces déterminations matérielles, je deviens donc un meilleur mécanicien et physiologiste de la situation en cours. Mais en même temps, je tisse un réseau d'associations dont les modes de parcours ne s'alignent en rien sur la succession des époques où mon imagination les inscrit.

Si donc je dispose de plusieurs perspectives sur la situation, selon ma façon de parcourir le temps à partir d'elle, et selon les variations virtuelles qui me permettent d'en construire le sens, le problème se pose peut-être de rechercher un accord, une correspondance, entre ces différents modes de saisie. Je peux décrire certains éléments de la scène selon un mode physique (la trajectoire du verre lorsqu'il s'est renversé), selon un mode intentionnel et physique (comment je l'ai renversé en voulant le saisir), selon un mode fonctionnel (ce qui se passe quand tout tourne rond sur le plan de travail), etc. Et rien ne prouve que multiplier les ontologies soit la bonne méthode pour assurer les passages, la coordination entre ces différentes perspectives. Une fonction, telle que je l'attribue à une structure présente, n'appartient cependant pas au présent de la scène. Ce n'est pas un ingrédient de ce présent, ou une propriété de la chose. *Fonction* est parfois le nom que nous donnons à une place dans une interaction matérielle, mais seulement si elle est ressaisie dans un réseau d'associations, dans une perspective qui n'est pas seulement physique. Et même si la perspective est principalement physique, il se peut qu'elle intègre une vaste circulation de causes et d'effets, ainsi qu'une vue singulière sur les formes et trajectoires observées (ainsi par exemple des *verres de vin* et des *verres à vin*, qui fonctionnent différemment, même quand ils sont pleins tous les deux). La fonction n'est donc pas un élément matériellement déterminant dans le présent des situations où nous la faisons jouer, en dépit de tous ses corrélats physiques. En toute rigueur, nous ne devrions pas dire que la structure *a* une fonction, mais plutôt que la structure *fait signe* vers un fonctionnement, tout comme le nom de la fonction, la phrase qui dit son rôle, fait signe à son tour vers différents modes de parcours des phénomènes. Attribuer une fonction ou un rôle permet justement d'attirer l'attention sur tel ou tel aspect de la matérialité du phénomène, de se rattacher à une gamme de situations orientant notre conduite, de raisonner à leur propos, etc. On objectera que la fonction se résout parfois, dans le cas des artefacts notamment, en la conservation d'un bilan matériel local, et que les parcours sémiotiques de la perspective fonctionnelle se spatialisent alors intégralement dans l'architecture du dispositif. Je répondrai à cela que notre perception inscrit précisément dans notre présent tout autre chose que de la présence locale pure et

simple, et qu'on ne comprendra jamais ce qu'est un artefact si on ne l'admet pas au préalable. L'être humain spatialise le possible, et décèle la possibilité dans l'espace, et c'est précisément dans le jeu entre la perception de la présence et celle du possible que s'ouvre le *fonctionnement sémiotique*. Les notions de finalité ou de fonction renvoient à des modes de parcours sémiotiques, qui font évidemment un usage constant de schèmes causaux, mais sans jamais s'y réduire. Dans le cas des êtres vivants, la finalité interne que nous leur prêtons a certes des contreparties du côté du phénomène observé, elle se détaille bien en causes et en effets. Son répondant dans la structure du phénomène est si fort qu'on est prêt à lui concéder l'objet lui-même, et non pas seulement notre façon de l'examiner. Cependant, ainsi que le dit Emmanuel Kant, cet objet qu'elle organise, “ elle ne le détermine pas ”.

Des considérations analogues valent pour les modèles que nous construisons. En tant que concepts et schèmes mathématiques ou logico-informatiques, ils ont tout leur rôle à jouer, à condition de pas les prendre eux-mêmes pour les phénomènes, de croire que parce qu'ils en représenteraient suffisamment et provisoirement la forme, ils en auraient la générativité singulière. À condition, donc, de ne pas s'en servir pour destituer des phénomènes qu'ils ne peuvent qu'aider à déterminer, et non constituer. En tant qu'idées de la raison scientifique, qui poursuit en eux un certain type de systématité, les modèles sont formidablement éclairants. Mais il ne faut pas en faire un usage constituant, c'est-à-dire croire que se détermine directement en eux seuls, ou se ramène à eux seuls, ou s'arbitre par eux seuls une expérience objective. S'y référer ne signifie donc pas que serait résorbé tout conflit d'interprétation, toute crise du sens pour une pratique scientifique. Laisser croire cela, c'est jouer un rationalisme étroit contre la raison<sup>1</sup>.

6. *Pour une notion de champ*. — Mais nous pourrions aussi, plutôt que de tenter de lutter sur le plan philosophique (combat pour lequel nous ne sommes pas fait), envisager les choses sous un angle plus pragmatique. Après tout, nous utilisons couramment ces concepts de forme, de fonction, de valeur, et pas seulement lorsque nous essayons de nous situer par rapport aux divers fonctionnalismes. Nous pourrions donc accepter

---

<sup>1</sup> Je ne fais pas tant allusion ici aux fonctionnalismes, qu'à l'objectivisme ambiant – avec lequel ils ont toutefois partie liée. La première sagesse serait peut-être de comprendre que les modèles ne sont pas irréductiblement liés aux épistémologies qui les enrôlent, qu'il y a donc toujours une autre façon d'utiliser la logique, les réseaux connexionnistes, ou les systèmes dynamiques, comme il y a souvent de nouveaux sens à donner à des modélisations anciennes.



une certaine fatalité du langage fonctionnaliste dans les recherches cognitives, à condition cependant d'en diminuer ou retarder les effets illusoires, d'en désamorcer la charge propre. En surface, nous reprendrions ce langage à notre compte, mais en le vidant de son sens. Il suffirait peut-être de conditionner le champ de travail, en l'éclairant en permanence d'un certain nombre de questions vives. L'insistance des questions suffirait, peut-on imaginer, à décourager les tentatives de contrôle (philosophiques ou autres). En voici quelques-unes :

- comment s'obtient, et s'entretient, la séparation entre l'intérieur et l'extérieur du système ?

- y a-t-il un rapport constituant, intrinsèque, entre le temps, l'espace et les formes ? Comment évolue la formalité des formes ?

- comment se détermine l'individuation et la segmentation des formes ? le système lui-même est-il une forme ?

- comment détermine-t-on la réitérabilité des formes, comment les catégorise-t-on ?

- la finitude de la machine est-elle constituante : de son temps, de ses formes, de ses fins possibles (faillibilité) ?

- à quoi le système est-il ouvert ? le possible (formes et valeurs) forme-t-il un ensemble déterminé, donné d'avance, ou un ouvert constructible ?

- y a-t-il détermination réciproque entre les formes, le mouvement, l'action, et l'affect ou émotion ?

- valeur et forme : toute la valeur s'épuise-t-elle dans un jeu de formes actualisables, la valeur n'est-elle que la forme de la valeur ? ou plus subtilement la valeur est-elle la possibilité ouverte par la forme qui la supporte provisoirement, en attendant la forme suivante qui réélabore ce possible ? la formalité des formes dépend-elle de la valeur, de la poursuite de fins valuées ?

- valeur et extériorité : y a-t-il un monde propre, au sens où il y va de l'existence du système de réussir à construire/s'approprier son environnement, tout en se constituant lui-même dans cette rencontre ? quel est le type de stabilité de ce monde, comment sa temporalité s'imbrique-t-elle à celle du système ?

- valeur et finalité : comment se détermine le rapport entre la finalité générale que l'observateur prête au système, et les fins particulières qui la spécifient ? quelles sont les fins vitales pour l'existence, et comment se comprend la réouverture constante de la finalité qui rend ces fins possibles ?

- valeur et matière : la valeur est-elle plus radicalement une possibilité ouverte par la matière, telle que nous la signifie vaguement la forme qui la

marque ? la machine elle-même (c'est-à-dire le domaine présumé d'un certain format physique caractérisant, pour l'observateur, une finalité sans fins préassignées) dépend-elle de la poursuite de fins valuées, pour réussir à se maintenir comme réserve systématique de marques distinctives ?

Une réponse positive à cette dernière question implique de sortir du fonctionnalisme, même méthodologique, et d'abandonner le concept de machine comme *base* explicative du cognitif. Mais on pourrait néanmoins conserver l'idée de machine comme format émergent, dynamique fondamentale lente du système. Ce format représenterait, relativement à l'observateur, le fond provisoire dont se détachent les formes scrutées par lui.

Peut-on aller plus loin, et esquisser une notion de *champ* déjà favorablement située relativement à ce questionnaire ? Voici en tout cas quelques slogans :

— *Il y a toujours plus de matière qu'on ne croit.* Ce qui veut dire d'abord que les formes ne pourront jamais suffire ; ensuite que la matière entre et sort, et donc n'est jamais la même ; enfin qu'il n'y a pas d'espace des états internes, puisque sa dimension devrait toujours changer.

— *L'individuation est toujours un processus.* Ce qui veut dire que les formes ne sont jamais que des processus, des produits de processus de segmentation spatiale et temporelle déterminant le flot de marques où les formes s'inscrivent. Rien de ce qui est intérieur au système n'est donné une fois pour toutes, mais tout doit être reproduit. En conséquence, l'individuation du système est toujours en cours, comme finalement celle de toutes les entités paraissant dans le champ, en tant qu'elles sont reliées à lui. Les lois sont aussi des lois du champ, et non des lois portant sur des formes déjà individuées.

— *Il y a toujours plus de modes d'organisation qu'on ne croit.* Ce qui veut dire qu'ils ne s'étagent pas nécessairement en niveaux, et peuvent rester hétérogènes.

— *L'intérieur est toujours déjà jeté dans son extérieur.* Ce qui signifie évidemment leur détermination réciproque, notamment par le mouvement ou action. Ce qui signifie encore que rien n'est cognitif qui ne soit inscrit dans cette co-détermination. Connaître veut dire constituer comme son propre l'extériorité dont on hérite, et se constituer soi-même à travers cette appropriation. *Jeté* signifie qu'être-au-monde implique puissance immédiate d'agir, impossibilité de ne pas agir. Le système n'a donc pas d'intentions, il est une intentionnalité.

On pourrait facilement prolonger cette liste, mais là n'est pas notre propos. On a seulement voulu ici, comme partout ailleurs dans ce texte, prendre au sérieux les prétentions “matérialistes” et naturalisantes des fonctionnalismes, et les déborder sous le poids de ce qu'il leur faudrait laisser entrer pour répondre à ces prétentions. On aurait pu tout aussi bien mener l'attaque à partir de la culture<sup>1</sup>, plutôt que de la nature, en reprenant d'ailleurs une bonne partie du parcours argumentatif effectué ici. La solitude du couple sujet-objet, l'oubli du tiers qui nous constitue dans le langage, le positivisme des faits qui dispense de les négocier au sein d'une culture, font en effet plutôt bon ménage avec la perspective fonctionnaliste et la clôture actuelle des sciences cognitives. Et les fonctionnalismes jouent là, finalement, un rôle positif, en nous permettant de comprendre toujours davantage, à partir de la critique que nous en faisons, à quel point la naturalisation et la culturalisation des recherches cognitives sont profondément solidaires.

7. *Polémiques.* — Peut-on totalement passer sous silence le fait que les fonctionnalismes relèvent du genre, ou réseau de genres, de la philosophie analytique ? Je ne crois pas. J'admire la structure de débats de la philosophie analytique prise en général, même si je désapprouve l'ostracisme — ou à tout le moins la relégation aux marges — dont est frappée, en son sein, la référence à la philosophie continentale. Car si débat il y a vraiment, Kant, Bergson, Husserl, Cassirer, Heidegger, Merleau-Ponty, parmi d'autres, devraient y être rattachés, et ce n'est pas se montrer exagérément élitiste que de demander ce que l'on gagne à fréquenter exclusivement Fodor, Dretske, Block, ou même l'agréable Dennett<sup>2</sup>. Bien sûr, il y eut un temps où la philosophie analytique était aussi continentale, mais apparemment il est passé — je veux dire que le désir en est passé, tout comme s'est perdu le sentiment d'une unité de l'interpellation philosophique à travers les siècles. Ainsi les auteurs

---

<sup>1</sup> A partir de la culture ou de perspectives parentes comme le collectif, le social, etc. (traces et architectures, territoires, rôles, hiérarchies, etc.).

<sup>2</sup> Toutefois, la référence à Husserl n'a jamais été complètement occultée, et des styles philosophiques “intermédiaires” sont apparus dans ce passage. Il y eut naguère les travaux de Follesdal, McIntyre, Dreyfus, ou ceux de Mulligan, Smith, et Simmons sur la Gestalt et la méréologie. Jean Petitot, de son côté, a de longue date inscrit son travail dans la lignée de Kant et Husserl, et en l'articulant à une mathématisation, créé un style philosophique et scientifique original, une philosophie naturelle de l'esprit à la fois cognitive et non analytique. Les relations de l'entreprise cognitive avec le continent husserlien se sont récemment intensifiées en France, avec plusieurs séminaires sur phénoménologie, psychologie, et naturalisation (et un colloque international organisé en octobre 1995 à Bordeaux par J. M. Roy).

comme Putnam, qui dénoncent cet état de fait, ne semblent guère écoutés<sup>1</sup>. Ceux qui, comme E. Tugendhat ou V. Descombes, reprennent les problématiques continentales dans le style analytique font figure d'originaux. Même de grandes oeuvres inclassables comme celles de Peirce ou Wittgenstein sont tenues en lisière de la réflexion cognitive. Et qui enfin, dans la tradition analytique actuelle, intègre explicitement à son travail philosophique des lectures aussi finement critiques et documentées que celles que Merleau-Ponty, par exemple, faisait dans ses cours à la Sorbonne ? La philosophie analytique est très loin d'être une philosophie du génitif, pour reprendre ici l'expression de S. Auroux, c'est-à-dire aussi une philosophie des disciplines constituées, de leur histoire, et de leur pratique effective, qu'il ne serait pas inutile de mieux connaître<sup>2</sup>.

D'un autre côté, la philosophie continentale ne peut s'estimer dépossédée de ce qu'elle a laissé échapper, et son mépris relatif, toujours sensible, pour les questions cognitives, la prive d'initiative. Mais je ne vois pas pourquoi nous devrions nous contenter du choix médiocre qui nous est laissé entre le catholicisme figé de la tradition continentale et le protestantisme amnésique de la philosophie analytique contemporaine. À nous de créer une scène philosophique et scientifique à la fois pluraliste et polémique, puisqu'il n'y a pas, comme le dit Emmanuel Lévinas dans un autre contexte qui vaut modèle pour le nôtre, de vérité sans la pluralité polémique des voix. "Polémique" implique en particulier d'organiser parfois la discussion autour de papiers-cibles ou *target papers* ; implique aussi d'introduire dans l'usage français (européen ?) l'absence de précaution et la fameuse franchise de discussion qui font souvent le charme des exposés outre-atlantique. "Pluraliste" veut dire le refus radical de toute espèce d'exclusion sectaire, du type de celles qui font la honte de certaines revues pourtant influentes. Pluraliste veut dire aussi la défense de tous les styles philosophiques et scientifiques, et donc l'éclosion de nouveaux genres de textes, qui ne seront pas tous du type argumentatif analytique, et intégreront aussi d'autres technicités que celle de la logique. Car enfin, le débat a beau être ouvert et interminable, débattre veut toujours dire parler sous l'obligation de conclure, et donc de se battre pour parvenir à la conclusion. Or en sciences comme en philosophies, nous concluons parfois, mais nous décidons beaucoup plus encore. Le profane ne peut que sourire ou s'exaspérer des enfantillages de

---

<sup>1</sup> Cf. l'entretien de H. Putnam avec C. Bouchindhomme dans *Définitions – pourquoi on ne peut pas "naturaliser" la raison* (Editions de l'Eclat, 1992, p. 48 notamment). Un "syndicat", dit-il aimablement de ses confrères...

<sup>2</sup> La logique mise à part, bien évidemment. A moins qu'il ne faille dire plutôt : *même* la logique.

la philosophie analytique, quand elle prétend clore et répondre, tuer définitivement un argument, au lieu de s'employer à réouvrir les questions et remettre à neuf le langage. Ma génération est ainsi passée du dogmatisme de la "théorie" en sciences humaines dans les années 60 et 70, aux programmes de recherche imaginaires de la philosophie de l'esprit, qui voudrait souvent qu'on la prenne pour une philosophie organique du laboratoire<sup>1</sup>. Ainsi le statut des vérités construites par les sciences du cognitif, leur clair-obscur, ce qu'elles nous font perdre ou acquérir, reste impensé. Mais pour cela il faudrait du temps, et la philosophie de l'esprit est trop occupée à poser elle-même les questions scientifiques.

On a donc compris que le *Schulbegriff*, le concept d'école de la philosophie analytique, ne me donnait pas satisfaction dans sa possible relation aux sciences cognitives (qui est ici ma seule préoccupation). Mais son *Weltbegriff*, sa forme de participation concrète au monde (par exemple celui de la recherche), n'est pas merveilleux non plus. Quelle que soit l'intelligence des personnes concernées, et le respect, voire l'obséquiosité avec laquelle les philosophes de cette école sont maintenant écoutés, du moins quand ils sont américains, je n'ai encore jamais rencontré un seul non-philosophe qui m'ait dit, avec le recul des années, avoir fait un pas décisif, connu une réorientation importante de son champ disciplinaire, ou tout simplement mieux compris ce qu'il faisait à partir de cette interaction. Le statut de la modélisation, en particulier, n'a jamais été correctement problématisé à partir de ce qu'elle est la plupart du temps : rapprochement analogique, métaphore et non simple procès de naturalisation, signe d'une corrélation des disciplines, plutôt que de l'unité organique de leurs objets<sup>2</sup>. Et la vérité, qui existe malgré tout, s'est bien vengée, car le malaise s'est maintenant installé, et personne ne veut plus s'en laisser conter par les autres. C'est toute l'interdiscipline cognitive qui en est lourdement marquée.

---

<sup>1</sup> Ce qui impliquerait qu'on l'écoute et l'évalue dans le même horizon temporel que celui des affaires du laboratoire, sous le même régime de significations et de preuves, etc. Trop souvent, la philosophie de l'esprit, notamment dans ses guises fonctionnalistes, refuse d'assumer une position de discours clairement extérieure au champ des sciences, si bien qu'elle s'y inclut sur le mode imaginaire. Assumer l'extériorité, pourtant, n'empêcherait pas l'accès à l'information scientifique la plus pointue. Bien au contraire, cette information n'en deviendrait que plus indispensable.

<sup>2</sup> On pourrait aussi évoquer les synergies négatives avec la psychologie cognitive et l'intelligence artificielle, mais il nous faudrait alors, sans noblesse aucune, céder à la tentation du pamphlet. Mieux vaut donc s'abstenir, hélas.

Mais après tout, ce n'est peut-être pas seulement le rôle d'une philosophie que d'authentifier les pratiques de la science ou de rendre leur cohérence lisible, contrairement à ce que semble aussi croire la *philosophy of mind*. C'est la crise et la vigilance critique qu'il faudrait plutôt lui demander d'apporter, et non l'apaisement des consciences. Mais voilà, cette philosophie semble totalement impuissante, et d'ailleurs privée du désir de nous aider à réfléchir au sens de notre entreprise de connaissance, à la qualité de nos projets, à la responsabilité qu'ils impliquent. Je me demande même si, à partir de ses ressources propres, et sans contrevenir à l'image de la connaissance qu'elle soutient, une telle philosophie serait encore capable de justifier sa propre structure de débat, encore une fois si admirable, si précieuse. On ira donc jusqu'à questionner l'évidence, et demander de qui elle tient cette structure, et pourquoi finalement elle s'y tient.

Yves-Marie VISETTI  
Laforia-CNRS, Université Pierre et Marie Curie  
4, Place Jussieu, 75252 Paris Cedex 05  
visetti@laforia.ibp.fr