

Bernard FELTZ\*

## Les paradoxes liés au concept de « modèle interne ».

*Résumé* : Les relations entre « modèle interne » et « représentation » sont analysées en lien avec la problématique du « théâtre cartésien » et la métaphore des « homoncles emboîtés ». Deux modélisations particulières sont étudiées : la représentation neurale du monde social de P. Churchland et la théorie de la sélection des groupes neuronaux de G. Edelman. L'analyse épistémologique porte sur les concepts de modèle et de représentation dans chaque démarche, en lien avec l'épistémologie de la modélisation mathématique en biologie, et en relation avec les travaux sur la robotique. Un rapprochement est finalement proposé avec les perspectives phénoménologiques de Merleau-Ponty.

*Mots-clés* : auto-organisation, concept, connexionnisme, conscience, donation de sens, émotions, EMPATH, groupes neuronaux, homoncule, instruction, langage, modèle, modélisation, modèle interne, naturalisation, phénoménologie, répertoires, représentation, réseaux neuronaux, robotique, sélection, sémantique, théâtre cartésien.

*Abstract*: **Some paradoxes of the concept: “internal model”.** This paper analyses the relations between “internal model” and “representation”, specially with reference to both “Cartesian theatre” and “endless regression of homunculi” problematics. Two specific models are studied: the neural representation of social world by P. Churchland and the theory of neuronal group selection by G. Edelman. Epistemological stakes concern the concepts of model and of representation in each approach, in relation with epistemology of mathematical modelling in biology, as with robotics. We finally take up the relations with Merleau-Ponty's work in phenomenology.

*Key Words*: Cartesian theatre, concept, connexionism, consciousness, emotions, EMPATH, homunculus, instruction, internal model, language, meaning giving, model, modelization, naturalization, neuronal groups, neuronal webs, phenomenology, repertories, representation, robotics, selection, self-organization, semantics.

---

\* Université catholique de Louvain, Centre de Recherche en Neurosciences, Institut Supérieur de Philosophie.

### INTRODUCTION.

La problématique du « Théâtre cartésien », pour reprendre l'expression de Daniel Dennett, ou des « homoncules emboîtés », selon la formulation de Gerald Edelman, est au cœur des questions épistémologiques liées au concept de « modèle interne ».

Pour Dennett, en effet, « le Théâtre cartésien est une représentation métaphorique de la façon dont l'expérience consciente doit avoir son siège dans le cerveau... Le matérialisme cartésien est la thèse selon laquelle il existe une ligne d'arrivée cruciale ou une frontière quelque part dans le cerveau, qui marque un endroit où l'ordre d'arrivée est identique à l'ordre de « présentation » dans le cerveau dans l'expérience parce que *ce qui se passe là* est ce dont vous êtes conscient. ».<sup>1</sup> Le dualisme cartésien des substances n'est plus de mise, mais la question des modalités d'interprétation des données de la conscience reste entière et la question du lieu de cette interprétation lui est associée, le plus souvent implicitement.

Edelman exprime cette même question à partir de la métaphore des « homoncules emboîtés ». « La notion d'instruction ou de traitement de l'information exige qu'il y ait quelqu'un ou quelque chose pour lire l'information. Mais une entité semblable s'avère alors nécessaire pour lire les messages résultants et ainsi de suite à l'infini. »<sup>2</sup> On entre dans la spirale de la régression à l'infini des instances interprétantes.

La double question des modalités d'élaboration du modèle et des procédures d'interprétation est au cœur de la problématique du « modèle interne ». D'autre part, cette double question peut être abordée différemment en fonction du type de modèle interne auquel il est fait référence.

C'est la raison pour laquelle je propose un exposé en trois parties. Dans les deux premières parties, je voudrais proposer une analyse épistémologique de deux modèles se référant à des formalismes et des logiques d'explication distinctes. L'objectif est d'analyser les modalités d'élaboration et d'interprétation du modèle... Comment, dans chacune des procédures, se pose la question du théâtre cartésien et des homoncules emboîtés, et comment chaque formalisme tente de résoudre ces questions. En une troisième partie, je proposerai une lecture plus philosophique des conclusions des deux premières parties, de la problématique de la modélisation de la conscience, en lien avec les travaux de Merleau-Ponty.

---

<sup>1</sup> Dennett, 1993, p. 141. Les soulignés sont de Dennett lui-même

<sup>2</sup> Edelman, 1992, p. 108.

## 1. UN MODÈLE CONNEXIONNISTE : P. CHURCHLAND ET LA REPRÉSENTATION NEURALE DU MONDE SOCIAL.

Dans un ouvrage intitulé *Minds and Morals*,<sup>3</sup> P. Churchland se réfère à un réseau de neurones soumis à un apprentissage pour reconnaître les émotions humaines. Je voudrais reprendre succinctement ce modèle pour proposer une analyse épistémologique de la manière dont se pose la question du théâtre cartésien dans les réseaux de neurones et expliciter quelque peu les ambiguïtés liées au concept de modélisation.

### 1.1. Approche scientifique

#### 1.1.1. EMPATH : un réseau pour reconnaître les émotions.

Le réseau de neurones imaginé par G. Cottrell et J. Metcalfe (California, San Diego) comporte trois niveaux. Le premier niveau consiste en un maillage de 64x64 neurones sensibles à la lumière ; chaque neurone comporte 256 niveaux de brillance. Le deuxième niveau comporte 80 cellules et constitue une dimension cachée du réseau. Le troisième niveau comporte 8 cellules, qui sont les cellules output ou cellules de reconnaissance ; à chaque cellule correspond une émotion. L'ensemble du réseau comporte donc  $(64 \times 64) + 80 + 8 = 4.184$  cellules avec 328.320 connections synaptiques

Ce réseau est entraîné par recours à 20 visages d'étudiants, dix garçons, dix filles ; pour chaque visage, 8 photographies représentent 8 émotions distinctes :

1 Étonnement, 2 joie, 3 plaisir, 4 détente, 5 somnolence, 6 ennui, 7 tristesse, 8 colère.

Le but est de voir si le réseau est capable de discriminer les caractéristiques à ce niveau de subtilité, sur une réelle diversité de faces humaines. Et il s'avère que la réponse est oui.

L'entraînement comporte 1000 présentations de 160 photos avec ajustement synaptique différentiel après chaque présentation. Après entraînement, le réseau est capable de reconnaître avec un haut niveau d'exactitude les 4 émotions positives (80%) ; par contre, il présente un bas niveau d'exactitude pour les émotions négatives, sauf pour la colère reconnue avec 85%.

De plus, il peut apprendre et généralise avec succès à des photos de visages qu'il n'a jamais vus auparavant.

Sa performance est robuste pour 5 des 8 émotions, et ses plus faibles performances sont corrélées avec les mêmes faibles performances chez des expérimentateurs humains. (Des tests ultérieurs sur des humains ont montré qu'ils présentent des difficultés à distinguer somnolence, ennui et tristesse). Cela signifie que reconnaître les 8 émotions est dans les possibilités d'un réseau de

---

<sup>3</sup> Churchland, 1995.

neurones et qu'un réseau plus étendu avec un entraînement plus long pourrait donner de meilleurs résultats.

Pour P. Churchland : « EMPATH » est « une preuve que, pour certains réseaux et pour certains comportements humains socialement signifiants, les premiers peuvent apprendre à reconnaître les seconds »<sup>4</sup>.

En fait, huit patrons d'activation différents sont produits au second niveau pour distinguer les huit émotions, et les trois patrons problématiques pour les émotions non reconnues sont diffus et indistincts.

#### *1.1.2. Complexification : caractères sociaux et séquences prototypiques*

Ce réseau est beaucoup trop simple pour rendre compte de ce qui se passe dans la réalité. Le réseau est construit sur base d'instantanés « hors temps ». Dans la vie de tous les jours, les émotions sont reconnues en fonction d'une situation plus large qui implique une séquence comportementale dans le temps. EMPATH n'a aucune notion des antécédents qui, typiquement, produisent les principales émotions, ni aucune notion des effets que ces émotions induisent sur le comportement cognitif, social et physique.

Une cognition sociale sophistiquée requiert la compréhension de patrons d'activation dans le temps et cela implique que le réseau soit enrichi par des voies récurrentes, voies additionnelles qui renvoient l'information de niveaux neuronaux supérieurs à des niveaux neuronaux inférieurs. Seule une telle structure permet la reconnaissance de séquences causales. De tels réseaux sont entraînés, non sur des instantanés, comme EMPATH, mais sur des séquences appropriées de patterns d'inputs.

Un sous-ensemble important des séquences causales consiste en des séquences conventionnelles ou rituelles qui font elles-mêmes l'objet d'apprentissage. Au terme de l'apprentissage, la bibliothèque des prototypes sociaux hiérarchiquement organisés dans le vaste espace d'activation neuronale de tout humain normalement socialisé doit rivaliser, voire dépasser la bibliothèque acquise des prototypes purement physiques.

En effet, en apprenant à représenter le monde, les cerveaux des enfants se focalisent naturellement sur les caractères sociaux de leur environnement local. Cela signifie que, lorsque les enfants partagent leurs espaces d'activation, les catégories que cela forme sont aussi souvent des catégories sociales que des catégories naturelles ou physiques. En répartissant des ressources neuronales pour des buts cognitifs importants, le cerveau s'étend bien au-delà de ces ressources

---

<sup>4</sup> Churchland, 1995, p.95.

en représentant et contrôlant la réalité sociale comme il le fait en représentant et contrôlant la réalité physique.

## 1.2. Aspects épistémologiques

Je voudrais tout d'abord préciser en préambule, que les perspectives proposées par P. Churchland me paraissent stimulantes sur le plan philosophique. Ses propos concernant les comportements prototypiques et sa conception aristotélicienne de l'éthique donnent à penser... même si elle prête à discussion. Mais venons-en à nos questions épistémologiques.

### 1.1.1. Théâtre cartésien – homoncules emboîtés

Implicitement à la notion de théâtre cartésien, on trouve la conception objectiviste de la perception qui était le schéma de l'intelligence artificielle dans sa phase classique :

Monde objectif  $\Rightarrow$  cerveau  $\Rightarrow$  sorties

Dans un tel schéma, la perception est conçue comme une sorte de miroir interne de la réalité extérieure. L'organisme est en situation de réception passive des entrées et comporte une instance interprétante centralisatrice. Un tel schéma comporte à la fois le « théâtre cartésien » comme lieu de centralisation de l'information et l'« homoncule emboîté » comme instance interprétante.

Par rapport à un tel schéma, la perspective connexionniste reste dans la logique de la perception « miroir », mais complexifie le modèle interprétatif. Il y a substitution de pattern de reconnaissance au traitement classique. Si, après apprentissage, le réseau est en mesure de reconnaître les diverses émotions sur les visages, c'est bien après un entraînement complètement contrôlé par une instance extérieure. Si, au niveau de la phase interprétative, l'homonculus est remplacé par la dynamique du réseau, cet homonculus a été renvoyé à l'extérieur pour la phase d'entraînement qui est absolument décisive pour le processus de reconnaissance. On reste donc dans un schéma où un programmeur externe finalise le réseau. A certains égards on explique une reconnaissance sur la base d'une instance qui reconnaît au préalable. Il y a donc circularité de la démarche. Si on évite partiellement le théâtre cartésien, on n'échappe pas à l'homoncule !

### 1.1.2. Le rapport à la modélisation.

Une autre objection est souvent faite à l'approche connexionniste concernant son lien à la réalité biologique en jeu dans les processus simulés.

L'analyse épistémologique de la modélisation de phénomènes biologiques montre qu'une phase importante dans la construction d'un modèle mathématique est le calibrage du modèle, autrement dit, la détermination des paramètres en fonction du phénomène spécifique à modéliser.

A titre d'exemple, j'ai suivi l'élaboration du modèle mathématique de l'évolution de la qualité de l'eau dans un bassin de *storage* par une équipe internationale. Les procédures et modèles mathématiques sont clairement définis. Il s'agit d'un système d'équations différentielles qui visent à simuler les différentes transformations de la matière organique dans l'écosystème aquatique : production primaire, *grazing*, activité détritique, échanges avec les sédiments... Toute une série de campagnes permettent de préciser, le plus souvent par voie indirecte, les divers paramètres des équations en présence. Pour la production primaire, par exemple, on distingue  $K_{max}$  ou taux maximal de productivité du phytoplancton,  $I_k$  intensité lumineuse au seuil de saturation... Ces paramètres sont des caractéristiques spécifiques du milieu particulier étudié : en l'occurrence le bassin de *storage*<sup>5</sup>.

Il faut bien constater qu'aucune procédure d'estimation précise des paramètres spécifiques d'une situation particulière n'est pratiquée dans les travaux connexionnistes. On n'est pas dans le registre de la modélisation mathématique au sens strict mais plutôt de la biologie théorique. La modélisation mathématique implique le « calage » du modèle sur une situation particulière : le modèle prend dès lors une capacité prédictive. En biologie théorique, le modèle consiste en un formalisme général qui tente de rendre compte de manière analogique des événements à décrire. La biologie théorique garde tout son intérêt heuristique, mais comporte un faible pouvoir prédictif et pose question sur le plan des mécanismes biologiques supposés.

La même question se rencontre dans la problématique de l'auto-organisation. Dans son ouvrage, *The Origins of Order. Self-Organization and Selection in Evolution*<sup>6</sup> St. Kauffman analyse le comportement de réseaux d'automates booléens au bord du chaos, et interprète ce comportement en termes de réalité biologique. A titre d'exemple, il considère le génome comme réseau d'automates booléens où chaque gène est un automate. Si, dans ce contexte, on considère une cellule différenciée comme un état stable du réseau, on peut rendre compte, de manière approchée, du nombre de cellules différenciées différentes, en fonction du nombre de gènes ; on peut rendre compte également du caractère arborescent du processus de différenciation cellulaire, associé à l'évolution d'un réseau d'automates à l'intérieur d'un même bassin attracteur.

Ici également, les interprétations sont à la fois stimulantes et problématiques. Stimulantes, car elles donnent à penser, et montrent que des phénomènes surprenants peuvent trouver une explication dans des dynamiques de systèmes complexes. Problématique, car le rapport à la réalité biologique pose question. En l'occurrence, le génome

---

<sup>5</sup> Pour une analyse détaillée, cf. Feltz, 1991.

<sup>6</sup> Kauffman, 1993.

humain fonctionne sur un modèle très différent d'un réseau d'automates 2x2.

C'est cette objection qu'Edelman développe en termes très différents dans sa postface à la biologie de la conscience pour justifier ses réticences au modèle connexionniste<sup>7</sup> et soutenir sa propre théorie que je vais maintenant présenter à titre de deuxième exemple.

## **2. EDELMAN ET LA THÉORIE DE LA SÉLECTION DES GROUPES NEURONAUX.**

### **2.1. Approche scientifique.**

L'objectif explicite d'Edelman est de proposer une anthropologie unitaire, non dualiste, sur base d'une analyse du fonctionnement du système nerveux. Il s'agit pour lui de « replacer l'esprit dans la nature » et de montrer comment on peut rendre compte de l'apparition de la conscience dans l'évolution biologique sans recourir à une substance différente du corps. C'est dans cette perspective que la théorie de la sélection des groupes neuronaux prend une place décisive.

La question centrale concerne les mécanismes de mise en place du système nerveux central des organismes supérieurs et de l'homme. La structure fine du système nerveux des organismes supérieurs est étroitement associée à l'activité même de ces organismes. C'est cette interaction complexe entre mise en place des tissus et activité de l'organisme qu'Edelman va appréhender au moyen de sa théorie.

#### *2.1.1. Instruction et sélection*

Pour aborder la problématique d'un point de vue très général, précisons d'emblée qu'Edelman se réfère à une logique explicative en termes de sélection et non d'instruction.

Dans une logique d'explication par instruction, la structure finale du système nerveux serait complètement définie par un programme détaillé qui préciserait l'ensemble des connexions à effectuer. On est donc dans le registre de la planification intégrale et d'un déterminisme absolu en fonction de cette préplanification. Dans le contexte de la biologie moléculaire, cette planification est associée au patrimoine génétique et à la structure de l'ADN qui en est porteuse.

Toute une série d'arguments théoriques et expérimentaux ont amené Edelman à élaborer une théorie qui se réfère non à une logique d'explication par instruction mais à une explication par sélection. Dans ce contexte, le programme ne précise pas la structure fine du système nerveux central : le détail des connexions synaptiques au niveau du cortex relève d'une autre logique. Le programme détermine l'apparition d'une structure hypercomplexe redondante, et c'est l'interaction avec l'environnement qui va sélectionner les connexions

---

<sup>7</sup> Edelman, 1992, p. 299.

pertinentes au détriment des autres connexions, qui vont dégénérer. Ce schéma sélectionniste, qu'Edelman reprend à la théorie darwinienne de l'évolution, est au cœur de la théorie de la sélection des groupes neuronaux et justifie son appellation de « darwinisme neuronal ».

Comment un tel schéma général s'applique-t-il concrètement au système nerveux ? C'est ce que nous allons analyser immédiatement.

### *2.1.2. Trois niveaux de sélection*

#### *a. répertoires primaires et sélection somatique*

L'organogenèse des structures de base du système nerveux central des organismes supérieurs se réfère à deux types de mécanismes. La mise en place d'une carte visuelle du toit optique chez la grenouille, par exemple, connaît deux phases. Une première phase génétiquement déterminée au sens strict, conduit à la formation d'une carte grossière. Une seconde étape conduit à un affinement de cette carte et est liée à l'activité des fibres nerveuses. Un certain nombre de prolongements axoniques émanent de la rétine et se dirigent vers le toit optique. Les terminaisons nerveuses qui aboutissent, de la sorte, à l'endroit correspondant du toit optique se stabilisent, tandis que les terminaisons nerveuses qui ne rencontrent pas le toit optique dégèrent. C'est un premier mécanisme de sélection appelé « sélection somatique ». Dès ce niveau fondamental, on n'a plus affaire à un déterminisme génétique strict, puisque la structure finale ne dépend pas seulement du programme, mais également du comportement de la cellule au cours du développement de l'organisme.

C'est précisément pour souligner l'importance du comportement cellulaire individuel qu'Edelman caractérise ce phénomène en termes de processus épigénétique dynamique, où la topologie, c'est-à-dire le rapport à l'espace, prend une importance considérable.

#### *b. répertoires secondaires et stabilisation sélective*

À un niveau plus complexe, le système visuel du singe par exemple comporte plus de trente-trois cartes visuelles, chacune spécialisée avec un certain degré de ségrégation fonctionnelle suivant l'orientation, la couleur, le mouvement... L'élaboration de telles cartes requiert un autre type de mécanisme. En effet, c'est le rapport à l'environnement et la stimulation des connexions nerveuses qui vont conduire à la stabilisation et au renforcement de ces connexions. C'est le processus de stabilisation sélective. Les connexions stimulées sont stabilisées tandis que les connexions non stimulées vont dégénérer.<sup>8</sup>

Un tel mécanisme présuppose par conséquent l'apparition, à un moment donné de l'organogenèse, d'un système hyperconnecté, à

---

<sup>8</sup> Ce schéma explicatif a été proposé par J.P. Changeux dès 1973. Cf. Changeux *et al.*, 1973. Cf. aussi Changeux et Danchin, 1974.

connectivité redondante, sur lequel une sélection va s'opérer en fonction du comportement de l'organisme sur la base du mécanisme de la stabilisation sélective. L'activité de l'organisme concourt donc de manière décisive à l'élaboration des répertoires secondaires du système visuel.

### *c. cartographie globale et réentrées*

Un troisième phénomène important est l'apparition des réentrées. En effet, les différentes cartes vont être reliées par des connexions réciproques de voies réentrantes, réseau très dense et redondant. Ces voies réentrantes conduisent à ce que la sélection de groupes neuronaux dans une carte induit la sélection simultanée d'autres groupes dans d'autres cartes. De telles connexions entre cartes peuvent donc expliquer la coordination d'activités d'un grand nombre de cartes.

Bien plus, si on prend en considération le fait que de telles connexions peuvent relier les diverses cartes interconnectées et les réseaux sensori-moteurs, c'est l'adaptation du comportement de l'organisme à l'environnement qui trouve là une source d'explication. En effet, si les processus d'apprentissage et de mémorisation sont liés à des processus de sélection de groupes neuronaux par stabilisation sélective, on comprend que l'apprentissage lui-même conduise à la stabilisation des connexions liées au comportement bien adapté. L'adaptation est donc le résultat d'un processus de sélection neuronale en fonction du comportement de l'animal.

#### *2.1.3. Deux niveaux de conscience*

Ces perspectives vont permettre à Edelman de proposer une hypothèse explicative de l'émergence de la conscience au cours de la phylogenèse. Pour cela, il distingue conscience primaire et conscience d'ordre supérieur. Par leur conscience primaire, les animaux supérieurs sont capables d'organiser leur comportement comme une réaction à l'environnement en fonction d'un mode de représentation de l'environnement caractérisé par l'immédiateté et la non-distance temporelle. Edelman parle à propos de la conscience primaire de « présent remémoré ». L'animal doté d'une conscience primaire est en quelque sorte « esclave du présent » dans la mesure où le recours à ses capacités de mémorisation sont dépendantes des stimuli de l'instant. Il vit donc dans l'instant et uniquement dans l'instant.

Par ailleurs, le langage joue un rôle essentiel dans le processus d'humanisation. Grâce au langage, l'être humain devient capable d'une représentation complètement déconnectée des stimuli de l'instant présent. Il devient donc capable de faire la distinction entre réel et imaginaire, entre passé, présent et avenir. La conscience d'ordre supérieur ouvre à une prise de conscience du temps, l'être humain est « conscient d'être conscient ». Le langage introduit une médiation importante entre l'individu et l'environnement qui lui donne une capacité de distanciation et d'élaboration de stratégies.

Conscience primaire et conscience d'ordre supérieur représentent donc des éléments importants dans la perspective darwinienne de la sélection naturelle. La logique sélectionniste permet de rendre compte de l'émergence de la conscience au cours du développement ontogénétique de l'individu. Cette même logique sélectionniste, au sens strictement darwinien cette fois, tente également de rendre compte de l'émergence de la conscience, primaire tout d'abord, de niveau supérieur ensuite, au long de l'évolution phylogénétique des diverses espèces au cours des temps géologiques. Les capacités, de représentation, tout d'abord, de distanciation, ensuite, ont manifestement constitué des avantages sélectifs tout au long de l'évolution.

## **2.2. Enjeux épistémologiques.**

### *2.2.1. Les homoncules emboîtés.*

Selon les dires d'Edelman lui-même, un des enjeux de sa théorie est de proposer un mécanisme qui respecte les propriétés biologiques du cerveau tout en échappant à la régression à l'infini de l'instance interprétative ou à l'externalité du programmeur.

La logique explicative sélectionniste, associée aux mécanismes biologiques de stabilisation sélective, dans des réseaux redondants reliant des cartographies partielles en une cartographie globale, permet d'expliquer l'émergence de réseaux synaptiques liés essentiellement au comportement et à l'histoire propre à chaque individu. La structure fine du cerveau est le produit de l'histoire de chacun.

Reste la question de la signification d'un comportement adéquat. Quels sont les critères qui spécifient un tel comportement ? Edelman parle à ce propos de « critères de valeur internes » qui ont été établis par sélection au cours de l'évolution. Ces valeurs se manifestent dans les régions du cerveau qui participent à la régulation des fonctions corporelles : rythme cardiaque, respiration, comportement sexuel, alimentaire, fonctions endocriniennes, végétatives... Ces valeurs imposent des contraintes aux domaines dans lesquels elles sont présentes. « Selon TSGN<sup>9</sup>, les forces motrices du comportement animal sont donc des ensembles particuliers de valeurs, sélectionnées au cours de l'évolution, qui aident le cerveau et le corps à maintenir les conditions nécessaires à la survie. Ces systèmes sont dits homéostatiques, et c'est le couplage du mouvement et de l'échantillonnage sensoriel, aboutissant à des comportements, qui modifie les niveaux des systèmes homéostatiques.»<sup>10</sup> La sélection naturelle au cours de la phylogenèse a donc retenu une série de « valeurs » qui servent de contraintes dans l'organisation du système nerveux et dans l'élaboration de comportements adéquats.

---

<sup>9</sup> Théorie de la Sélection de Groupes Neuronaux

<sup>10</sup> B.C. p.124.

A titre d'exemple, Edelman se réfère à des automates complexes élaborés par son équipe, en l'occurrence le Darwin III, où des valeurs telles que « la présence de la lumière est mieux que l'absence de lumière » et « la présence de lumière et la stimulation au centre du champ visuel sont préférables à la présence de lumière et la stimulation à la périphérie » conduisent l'œil de l'automate à suivre la trace des signaux provenant d'objets éclairés. Bien plus, il montre comment l'intégration de cartographies plus complexes intégrant plusieurs modalités sensorielles et articulées au système moteur conduisent l'automate à un comportement de catégorisation perceptive plus complexe.

Il apparaît par conséquent que, dans le contexte de la TSGN, l'adoption d'un « comportement adéquat » ne relève pas d'une programmation spécifique mais est un phénomène épigénétique lié, d'une part, à un système de valeurs retenues par la sélection au cours de la phylogenèse et, d'autre part, au processus de sélection de groupes neuronaux en fonction de l'expérience au cours de l'ontogenèse.<sup>11</sup>

Tout apprentissage comporte un rapport à la valeur. Sans renforcement lors d'une bonne réponse, pas d'apprentissage. D'où vient le renforcement ? Dans les réseaux neuronaux, c'est le programmeur et l'expérimentateur qui président à l'apprentissage. Dans le contexte de la théorie de la sélection des groupes neuronaux, c'est la sélection naturelle qui a retenu des valeurs qui s'imposent aux systèmes. On retrouve un schéma classique de la pensée biologique contemporaine qui recourt au concept de sélection naturelle pour rendre compte de l'organisation en fonction de la préservation du tout.

C'est une position très forte, qui permet d'échapper à la régression à l'infini de l'instance interprétante ou programmatrice, tout en proposant un système explicatif de l'adaptation du comportement. C'est une position qui laisse ouverte la question des modalités concrètes de sélection des valeurs.

Quoi qu'il en soit, il me semble que l'on peut dire que c'est une avancée très sérieuse par rapport aux modèles de réseaux neuronaux.<sup>12</sup>

---

<sup>11</sup> Les exemples de valeurs proposées par Edelman portent essentiellement sur les équilibres homéostatiques et touchent aux centres hédonistes dans l'apprentissage. Par ces exemples, essentiels du point de vue de l'évolution biologique, l'auteur ne prétend pas « épuiser » la problématique des valeurs dans le comportement humain. Cf. notamment les réflexions sur les valeurs au niveau anthropologique, p. 225.

<sup>12</sup> Une critique scientifique de la théorie de la sélection des groupes neuronaux dépasserait le cadre de cet article. Elle devrait impliquer à la fois une analyse de l'articulation entre explication par instruction et par sélection, une prise en compte des données récentes concernant la plasticité du système nerveux central, et les objections portant sur l'inscription de la dynamique sélectionniste dans un processus d'autonomisation.

### 2.2.2. *Le concept de représentation.*

Il est un autre aspect de la question où la théorie de la sélection des groupes neuronaux présentent une nouveauté sensible par rapport aux anciens modèles, cela concerne le concept de représentation. Je voudrais me référer aux travaux épistémologiques d'Andy Clark pour tenter de mieux cerner cette nouveauté.

Andy Clark<sup>13</sup> propose une réflexion sur l'évolution du concept de représentation. Sur la base de ses analyses des modélisations, en particulier de sa pratique de la construction de robots, il propose une évolution du concept de représentation.

Il parle d'une transition progressive de modèles de représentation comme miroir ou encodage à des modèles de représentation comme contrôle. Pour lui, le cerveau ne devrait pas être considéré comme fondamentalement un lieu de description des états externes du monde mais plutôt comme une structure interne qui agit comme un opérateur sur le monde, par son rôle de détermination de l'action.

Il prend l'exemple de cartes élaborées par des robots. Dans ce contexte, la carte joue à la fois la fonction de description du monde extérieur et de prescription de l'action. La perception est préparation de l'action. Les représentations internes sont mieux comprises comme des structures de contrôle de l'action et spécifiques à un contexte donné, que comme récapitulations passives de la réalité externe. Il plaide, dès lors, pour ce qu'il appelle une perspective émergentiste qui se centre sur les interactions entre corps, monde et cerveau.

« ...appelons l'histoire d'un processus représentationnel si elle dépeint des systèmes complets d'états internes identifiables (locaux ou distribués) ou des processus (des séquences temporelles de tels états) comme ayant la fonction de fournir des types spécifiques d'information sur un état des choses (*state of affairs*) externe ou corporel. »<sup>14</sup> Au niveau d'une théorie de la représentation, on peut parler d'un continuum. Le niveau le plus élémentaire serait la simple corrélation causale – la fleur qui se tourne vers le soleil ou le robot qui cherche la lumière – A. Clark parle d'un « relais adaptatif » (*adaptive hookup*). La composante représentationnelle prend de l'ampleur si on rencontre des états internes qui présentent une coordination systématique avec un espace complet de contingences environnementales. À l'autre extrémité de ce continuum, on trouve des états internes qui peuvent déployer des codes internes en l'absence totale des caractéristiques environnementales visées.

Il me paraît pertinent d'interpréter les perspectives edelmaniennes à la lumière des considérations épistémologiques de A. Clark. Chez Edelman, en effet, les cartographies secondaires et globales sont à la

---

<sup>13</sup> Clark, 1997.

<sup>14</sup> Clark, 1997, p. 147.

fois produit de l'histoire propre, conséquence du comportement passé, et préparation du comportement futur. On est donc bien dans le registre de la représentation comme médiation pour l'action. Toute la conception edelmanienne de la perception s'articule à l'action, à l'interaction de l'organisme et de l'environnement. On est donc bien dans le contexte d'une priorité de la relation de l'organisme avec le monde.

Même la dynamique de la conscience d'ordre supérieur est marquée par cette logique. En effet, dans son approche du langage, Edelman prend ses distances par rapport à la conception du langage de Chomsky qui donne priorité à la syntaxe et parle d'une grammaire universelle. Edelman au contraire donne une priorité au sémantique<sup>15</sup> et voit la grammaire comme le fruit d'une complexification progressive et évolutive du langage. Cette priorité du sémantique sur le syntaxique, du concept sur la grammaire, est explicitement liée à une approche pragmatique du langage. Le concept, conçu initialement dans sa fonction de dénotation, comme visée d'objet, est d'emblée un moyen de communication entre individus. L'évolution des espèces conduit à une complexification progressive et à une distanciation progressive de l'action. Mais à son origine, le langage est moyen de communication.

Tant sur le plan du concept de représentation que sur le plan des modalités d'interprétation, la théorie de la sélection des groupes neuronaux proposent donc des perspectives novatrices particulièrement stimulantes.

Par ailleurs, il me paraît également pertinent de rapprocher ces analyses de la problématique du modèle interne telle qu'elle est exposée dans les contributions scientifiques de cette journée. En effet, le continuum proposé par A. Clark à propos du concept de représentation me paraît adaptable au concept de modèle interne. L'usage de ce concept est en effet très large ; il renvoie aussi bien à la structuration du système nerveux lui-même pour rendre compte des réactions de l'organisme qu'à des modèles cognitivistes où se pose seulement en un deuxième temps la question de l'inscription dans la structure du système nerveux. Le continuum proposé par A. Clark trouve ici toute sa pertinence. Le même concept couvre un très large spectre qui va de la réaction liée à une simple boucle de rétroaction jusqu'aux élaborations complexes liées à la perception visuelle. Un travail d'explicitation des spécificités de chaque niveau de complexité du modèle interne et de son rapport à une inscription dans le système nerveux permettrait certainement de clarifier les débats et de lever bon

---

<sup>15</sup> Edelman recourt à une notion très large de « concept » : chez l'animal, le concept est une notion pré-langagière, une catégorisation émergente individuelle, qui influence le rapport à l'environnement. Le prédateur reconnaît sa proie, et la proie reconnaît le prédateur... Une telle conception peut être rapprochée de la fonction sémantique du langage, prise elle-même en un sens large.

nombre d'ambiguïtés. L'hypothèse d'un continuum me paraît très stimulante.

Pour un philosophe européen continental, cette insistance sur le rapport au corps et l'interaction avec l'environnement dans une perspective d'action rappelle notamment des tentatives déjà vues. Le rapport au monde est central dans toute la phénoménologie, et, dans ce courant, le rapport au corps a pris une place décisive essentiellement chez Merleau-Ponty. C'est pourquoi, je voudrais en une dernière partie, rapprocher la problématique de la naturalisation de la conscience, telle qu'Edelman la propose, avec les conclusions de *la Phénoménologie de la perception*.

### 3. NATURALISATION DE LA CONSCIENCE ET DONATION DE SENS

Merleau-Ponty se réfère à la problématique de la *Sinn-Gebung* husserlienne pour préciser les rapports entre les contraintes corporelles et la signification des comportements<sup>16</sup>. Il distingue trois attitudes. Dans la perspective réaliste, le sens est inscrit dans les contraintes corporelles. Le sens préexiste au sujet qui doit inscrire son comportement dans ce sens préexistant. Merleau-Ponty prend l'exemple du marxisme où le sens de l'histoire préexiste aux individus, qui ont à inscrire leur action dans ce sens prédéfini. Paradoxalement, certaines tendances théologiques recourent au concept de nature pour expliciter un sens prédéfini auquel les individus doivent se soumettre. Pour notre auteur, dans la perspective idéaliste, par contre, le sens est pure projection du sujet. C'est le sujet qui choisit le sens d'un comportement. Le corps est perçu comme obstacle à un sujet qui se perçoit comme tout puissant. Il n'y a pas de nature humaine ; l'homme est un être de culture. La perspective phénoménologique de Merleau-Ponty distingue un mouvement centripète et un mouvement centrifuge. Dans le mouvement centripète, le corps ouvre à des significations multiples à l'intérieur desquelles le sujet se projette dans un processus d'auto-détermination. Ainsi, pour Merleau-Ponty, l'insistance sur le corps et le rapport au monde ne conduit pas à nier la liberté mais à montrer comment la liberté s'inscrit dans un corps et dans un monde qui apparaissent comme condition de possibilité de cette liberté même.

De même, chez Edelman, le refus du dualisme cartésien et la naturalisation de la conscience correspondent à une volonté de réinscrire la conscience dans la nature. Mais ce faisant, il restitue à l'être humain une conscience intentionnelle qui situe l'être de culture comme produit évolutif de l'être de nature. Tout comme Merleau-Ponty, on peut dire qu'Edelman, à distance des perspectives réductionnistes radicales, opère une réconciliation à dimension

---

<sup>16</sup> Je me réfère plus particulièrement au chapitre conclusif de *la Phénoménologie de la perception* intitulé: Liberté. Cf. Merleau-Ponty, 1945, 496-520.

multiple de l'homme avec son corps, avec la nature, avec la science<sup>17</sup>.

### Références bibliographiques

- Changeux, J.P., Courrège, P., et Danchin, A., 1973. A theory of the epigenesis of neural networks by selective stabilization of synapses. *Proc.Nat. Acad. Sci. USA*, 70, 2974-2978.
- Changeux, J.P. et Danchin, A., 1974. Apprendre par stabilisation sélective de synapses en cours de développement, in Morin, E. et Piatelli, M., *L'unité de l'homme. 2. Le cerveau humain*, Seuil, Paris, 58-88.
- Churchland, P. 1995. The Neural Representation of the Social World, in May, L, Friedman, M. and Clark, A., Eds, *Minds and Morals. Essays in Cognitive Science and Ethics*, Bradford Books, MIT Press, p.91-108.
- Clark, A., 1997. *Being There. Putting Brain, Body, and World Together Again.*, Bradford Book, MIT Press, Cambridge, London.
- Dennett, D., 1993. *La conscience expliquée*, Odile Jacob, Paris.
- Edelman, G., 1989, *The Remembered Present : a Biological Theory of Consciousness*, Basic Book, New-York.
- Edelman, G., 1992. *Biologie de la conscience*, Odile Jacob, Paris.
- Edelman, G. et Tononi, G., 2000. *Comment la matière devient conscience*, Odile Jacob, Paris.
- Feltz, B., 1991. *Croisées biologiques. Systémique et analytique. Ecologie et biologie moléculaire en dialogue*. Ciaco, Erasme, Namur.
- Feltz, B., 2000. L'inné et l'acquis dans les neurosciences contemporaines, *Revue Philosophique de Louvain*, 98, 4, 711-731.
- Feltz, B., 2003. *La science et le vivant*, De Boeck, Bruxelles.
- Kauffman, St., 1993. *The Origins of Order*. Oxford University Press, New York, Oxford.
- Merleau-Ponty, M., 1945. *Phénoménologie de la perception*, Gallimard, Tell, Paris.

---

<sup>17</sup> Pour une analyse détaillée des relations entre Edelman et le réductionnisme, cf. Feltz, 2000. Le rapprochement entre Edelman et Merleau-Ponty est développé dans « Qu'est-ce que l'homme? Neurosciences, conscience, liberté » in Feltz, 2003, 161-199. L'objet de ce chapitre est de tenter une conciliation entre les positions non-dualiste et non-réductionniste.