

Jusqu'ici : on a "parlé cerveau" = **neurosciences**

= des méthodes pour étudier
la structure et le fonctionnement de
cet organe qu'est le **cerveau**, afin de
mieux comprendre la **pensée humaine**.



Mais **d'autres disciplines** essaient aussi de comprendre le
fonctionnement de nos facultés mentales ou "cognitives" en passant par
d'autres voies :

la psychologie,
la linguistique,
la philosophie,
l'informatique ("intelligence artificielle"),
etc.

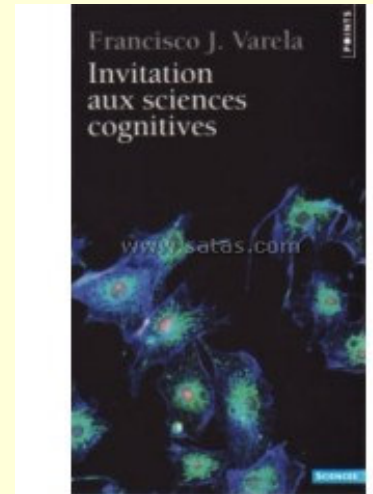
C'est de toutes ces autres démarches dont on va parler aujourd'hui.

On parlera donc peu de la vie de nos 4 neurobiologistes.
Mais **Francisco Varela** sera présent par l'entremise de son livre :

« Invitation aux sciences cognitives » (1996)

[la **réédition** mise à jour par Varela de
« Connaître les sciences cognitives. Tendances et perspectives. »
(1989) qui était la **traduction** française de
« Cognitive science. A cartography of current ideas. » (1988)]

qui m'a beaucoup inspiré pour préparer cette séance.



Autre source d'inspiration : Le cerveau à tous les niveaux !

le sujet 3 dans le thème

« L'émergence de la conscience »

qui s'intitule L'APPORT DES SCIENCES COGNITIVES



(page http://lecerveau.mcgill.ca/flash/d/d_12/d_12_p/d_12_p_con/d_12_p_con.html#3 et suivantes)

Bref, un cours un peu plus classique que d'habitude, pour tenter une introduction aux **sciences cognitives**

à travers trois grands **paradigmes** :



- le cognitivisme
- le connexionnisme
- les systèmes dynamiques incarnés

Ou comment terrasser un dragon à 3 têtes en 1 heure ? Finalement impossible...

[Donc : les 2 premiers aujourd'hui, le 3^e au prochain cours...]

Mais d'abord, qu'est-ce que les « **sciences cognitives** » ?

Et qu'est-ce qu'un « **paradigme** » en science ?

Qu'est-ce que les « **sciences cognitives** » ?

(terme pratiquement inexistant jusqu'au milieu des années 1970)

1970 : on assiste depuis des années à une convergence d'idées entre des psychologues, des informaticiens en « intelligence artificielle », des linguistes d'inspiration chomskienne, et quelques philosophes.

Mais on ne pouvait **pas encore parler d'un véritable discipline**, avec un paradigme unifié et une base institutionnelle solide.

A partir de 1975, les choses changent. Aux États-Unis, les « sciences cognitives » commencent à faire parler d'elles et s'organisent d'un triple point de vue : théorique, institutionnel, médiatique.

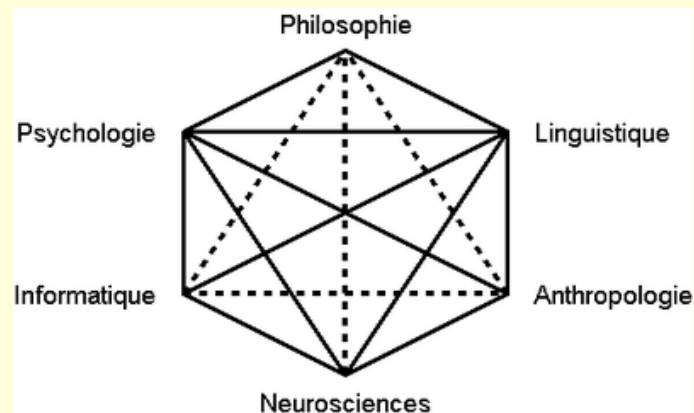
Théorique, avec un « paradigme cognitif » qui s'impose autour du modèle symbolique de Jerry Fodor, celui du traitement de l'information symbolique.

[on va reparler de tout cela en détail dans un instant...]

- **Institutionnel**, car c'est en **1975** que la fondation privée Alfred P. Sloan, va injecter 20 millions de dollars pour financer des recherches en « sciences cognitives » dans tous les États-Unis.

Elle finance aussi la revue, ***Cognitive Science***, dont le premier numéro paraît en **1977**, et la création d'une **société savante** en **1979**.

En **1978** paraît un premier rapport sur l'état des sciences cognitives, avec pour la première fois le fameux hexagone des disciplines concernées :



- **Médiatique**, car au début des années **1980** se multiplient les manuels, ouvrages d'introduction, livres de vulgarisation.

En **1985**, le psychologue Howard Gardner publie la première *Histoire de la révolution cognitive*, sous-titrée *Une nouvelle science de l'esprit*.

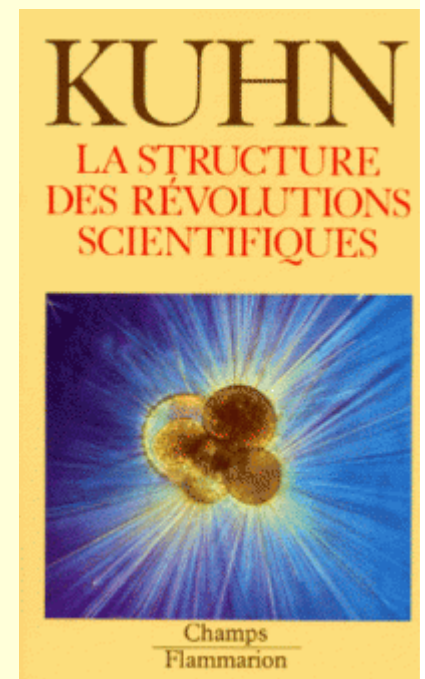
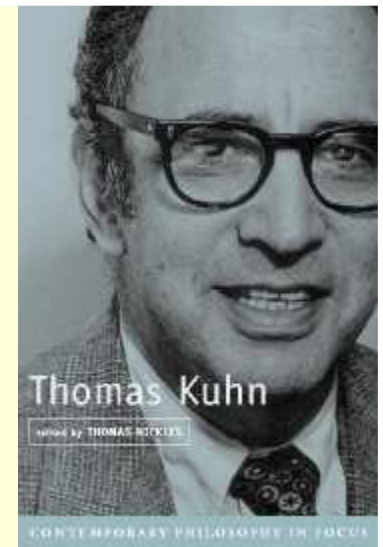
En **1988**, le « Cognitive science. A cartography of current ideas. » de Varela.

Qu'est-ce qu'un « **paradigme scientifique** » ?

Thomas Kuhn les a caractérisés dans son ouvrage ***La structure des révolutions scientifiques*** en **1962** comme un cadre de référence sur lequel s'appuient les chercheurs d'une discipline pour structurer et orienter leurs recherches à une époque donnée.

Par exemple, la médecine occidentale actuelle qui explique les troubles physiques par des causes physiologiques et les traite par des soins organiques (médicaments, opérations...) est le **paradigme dominant de la médecine officielle**.

Pour Kuhn, l'histoire des sciences est soumise aux contingences culturelles et sociales du moment et procéderait **par bonds** : à des périodes calmes où règne un **paradigme dominant** succèdent des **crises** de contestation pouvant déboucher sur des remises en cause radicales paradigmes du moment. Et parfois le remplacer...



Comme disait Jean-François Dortier dans un article publié dans le magazine Sciences Humaines :

« Comme toute révolution digne de ce nom, la « **révolution cognitive** » a renversé un ordre ancien, s'est emparée du pouvoir de façon exclusive, et a forgé une nouvelle orthodoxie. Mais son succès a aussi provoqué des réactions hostiles et des critiques sévères. »

Et c'est ce que l'on va constater en examinant les paradigmes du cognitivisme et du connexionnisme (cette semaine) et des systèmes dynamiques incarnés (à la prochaine séance).

IMPORTANT :

Ne pas confondre l'adjectif « cognitif » comme dans « sciences cognitives », dans le sens d'un ensemble de disciplines avec un programme commun;

avec l'adjectif « cognitiviste » ou le nom « cognitivisme » dans le sens d'un paradigme, c'est-à-dire d'une école de pensée particulière au sein des sciences cognitives.

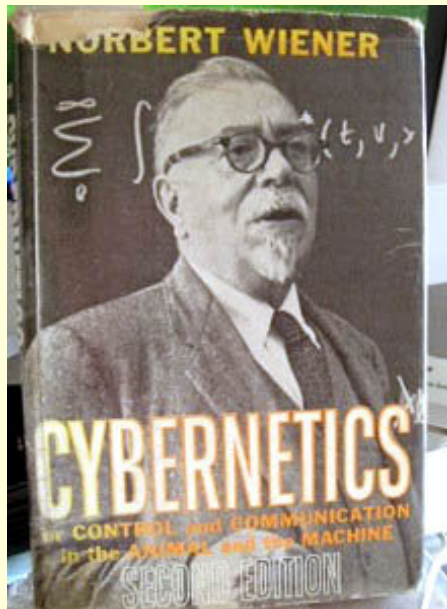
Vue d'ensemble des **3 paradigmes** en une diapo !

Le **cognitivism** domine les sciences cognitives du milieu des années 1950 aux années 1980. Il s'intéresse à l'esprit qu'il compare à un ordinateur.

Le **connexionnisme**, commence à remettre en question l'orthodoxie du cognitivism au début des années 1980. Il prend en compte le cerveau et essaie de comprendre la cognition avec des réseaux de neurones.

Les **systèmes dynamiques incarnés** vont critiquer les deux premiers paradigmes à partir du début des années 1990. Ils vont prendre en compte non seulement le cerveau, mais le corps particulier d'un organisme et l'environnement dans lequel il évolue en temps réel.

Et dans la recherche contemporaine, les trois approches coexistent, à la fois séparément et sous diverses formes hybrides.



On a vu au dernier cours [qui est sur le site de l'UPop] l'effort multidisciplinaire que fut **les conférences Macy** avec la réunion de plusieurs disciplines qui ont accouché d'idées nouvelles comme la cybernétique.

De même, ce qu'on va appeler les **sciences cognitives** peut être vu comme la reprise par des disciplines scientifiques contemporaines de questions philosophiques les très anciennes concernant **l'esprit humain**, son organisation, sa nature, les relations qu'il entretient avec le cerveau, avec l'organisme, avec autrui et avec le monde physique.

Il ne faut **pas voir ça** non plus comme une **succession stricte dans le temps**.

Il y avait déjà du connexionnisme dans les années **1940**.

L'idée d'énaction de Varela était déjà pressentie chez Merleau-Ponty début **1940**.

Des principes des systèmes dynamiques étaient déjà dans « An Introduction to Cybernetics », de Ross Ashby, en **1956**

(mais ça a pris plus tard les ordinateurs pour pouvoir calculer les attracteurs, etc.)

Mais c'est vrai que pour une **époque donnée**, il y avait un **paradigme dominant**.

Petit rappel historique pour mieux sauter dans les sciences cognitives...

Au XIXe siècle et au début du XXe, il existait une tradition de recherche en psychologie appelée **structuralisme**.

Représentés par des chercheurs comme Wilhelm Wundt et son élève Edward B. Titchener, les structuralistes utilisent la méthode expérimentale en psychologie, mais également **l'introspection** pour analyser les perceptions conscientes.

On voulait décrire les **composantes élémentaires** de l'esprit humain et essayer de comprendre de quelles façons toutes ses composantes s'incorporent ou **se « structurent »** en formes plus complexes.

Cette approche fut surtout critiquée pour la **difficulté de vérifier expérimentalement l'introspection** qui était à la base du structuralisme.

Ainsi naquit, à partir des années **1920**, un mouvement radicalement opposé à l'introspection du structuralisme : le **behaviorisme**.

Pour ses pionniers comme John B. Watson et B. F. Skinner, on ne pouvait bâtir une approche scientifique de la psychologie sur des états subjectifs de nature essentiellement **privée**.

Au contraire, cette nouvelle psychologie devait être basée uniquement sur **l'étude expérimentale du comportement** (d'où le nom qui vient de « behavior », comportement en anglais), et non plus sur les jugements individuels relatifs à nos états d'âme.

Seuls seront donc étudiés les **phénomènes observables**, c'est-à-dire les **stimuli** qui s'exercent sur l'organisme et les **réponses** que donne cet organisme.

Et l'on cherche à découvrir les lois qui gouvernent les relations entre les entrées et les sorties.



Cerveau = "boîte noire" = ce qui s'y passe est, par nature, méthodologiquement inaccessible et inobservable. [pas de scan !]

Un de leur **champ de recherche favori** était **l'apprentissage associatif** (l'enfant qui apprend à ne pas pas toucher une flamme car elle brûle...).

Et les behavioristes firent plusieurs découvertes, en particulier sur le **conditionnement opérant** des comportements, en expérimentant avec des pigeons, des rats et des chiens.

Il s'agissait donc d'une école de pensée extrêmement centrée sur l'influence de l'environnement sur nos processus mentaux.

Watson allait même jusqu'à dire que la structure de notre esprit est **entièrement façonnée par les récompenses et les punitions de notre environnement**, et par aucune influence génétique.

D'où cette blague de ses détracteurs qui faisaient remarquer qu'un behavioriste qui en rencontre un autre n'aurait pas d'autres choix que de lui dire :
« Vous allez bien aujourd'hui !
Et moi, comment vais-je ? »...



Pour les behavioristes, la science du comportement doit utiliser une méthode scientifique rigoureuse, basée sur des observations utilisables et quantifiables par tout chercheur.

D'où leur refus de toute spéculation sur des états mentaux qui n'étaient pas directement visibles par l'expérimentateur.

Les behavioristes vont donc bannir des sujets d'étude comme **la pensée, l'esprit, la conscience ou l'imagination,**

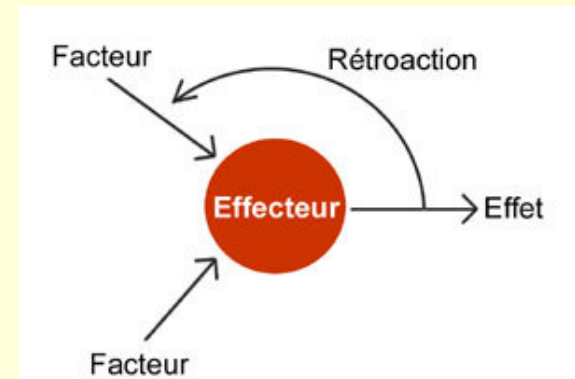
et des constructions hypothétiques comme **les symboles, les idées ou les images mentales.**

Ça fait beaucoup dans la poubelle...

Arrive alors la fameuse série de conférences Macy de 1946 à 1953, et de ce courant de pensée qu'elle allait générer : la **cybernétique**.

La cybernétique s'intéressait à la façon dont circule l'information tant chez les êtres vivants que dans les systemes complexes artificiels conçus par l'être humain.

Et on a vu que ce qui empêche la première cybernétique de se réduire à un simple **behaviorisme** obéissant au schéma stimulus-réponse, c'est la **notion de feedback** qui permet de changer la relation qu'établit un sujet entre input et output, entre stimulus et réponse.



Par ailleurs, la cybernétique, contrairement au behaviorisme, prenait en compte les données sur le cerveau.

Elle va aussi utiliser les deux approches des systèmes complexes:

- l'approche par réduction, **du haut vers le bas** («**top down**», en anglais);
- l'approche de construction globaliste ou systémique, **du bas vers le haut** («**bottom up**», en anglais).

Ces deux approches, plutôt complémentaires que contradictoires, vont d'ailleurs être derrière les deux premiers paradigmes des sciences cognitives dont on va bientôt parler :

le cognitivism (**top down**) et le connexionnisme (**bottom up**).

Parallèlement, un premier coup fut porté au behaviorisme dès **1948**.

Le **psychologue Karl Lashley** affirme alors que la conception du comportement sous forme d'arcs réflexes, du type A entraîne B, **ne permettait pas selon lui de décrire la complexité de certains comportements humains**, par exemple, la capacité de parler ou de jouer d'un instrument de musique.

Vu la rapidité avec laquelle un pianiste enchaîne les notes d'un morceau, on ne peut imaginer que la frappe de la première touche entraîne celle de la suivante. Il faut obligatoirement supposer qu'il y a une **anticipation** du comportement, dit Lashley.

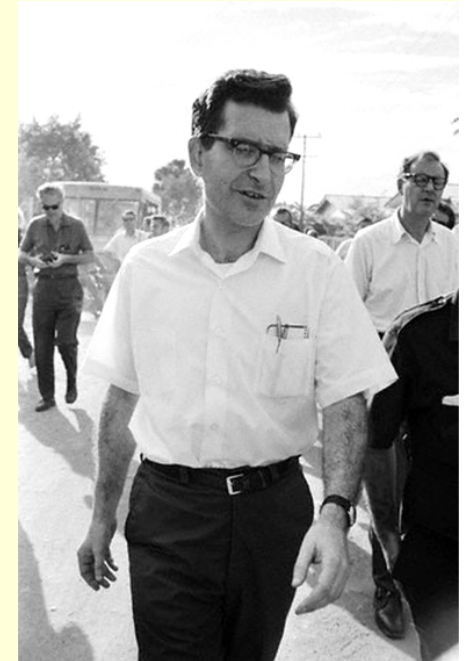
Il introduisait ainsi la notion, en totale opposition aux théories béhavioristes, que l'organisation du comportement **émane de l'intérieur de l'organisme.**

Dans son *Histoire de la révolution cognitive*, Gardner considère cette conférence de K. Lashley comme un des **éléments déclencheurs de la révolution cognitive.**

À cette époque se développait aussi la **linguistique** en tant que véritable domaine scientifique consacré à l'une de nos capacités mentales les plus sophistiquées, **le langage**.

Une des critiques les plus sévères du béhaviorisme va venir de celle que le linguiste **Noam Chomsky** adresse en **1959** à Skinner où il affirme que vouloir étendre le modèle béhavioriste de l'apprentissage à la linguistique est **sans espoir**.

Pour lui, les compétences linguistiques ne peuvent, même en principe, être expliquées sans admettre que les êtres humains possèdent un répertoire relativement important de **structures cognitives complexes** qui président à l'usage du langage.



L'esprit humain était donc de moins en moins considéré comme une boîte noire et toutes ces nouvelles disciplines (cybernétique, linguistique, informatique, etc) commencèrent de fructueuses interactions qui allaient mener aux « sciences cognitives ».

Celle-ci allait garder l'exigence de rigueur scientifique des behavioristes, mais se donnait comme objectif **d'identifier et de comprendre les états mentaux d'un individu.**

On parle souvent de **1956** comme de l'année où naquit la psychologie dite « cognitive » grâce, entre autres, à deux professeurs de Harvard, **Jerome S. Bruner** et **George Miller**.





Bruner travaillait à l'époque sur le processus de « **catégorisation** ».

En demandant à ses étudiants de classer des cartes de couleurs et de formes différentes, Bruner s'était aperçu que les individus utilisaient des **stratégies mentales différentes**. Les uns procédaient à partir d'une carte de référence, d'autres effectuaient un classement fondé sur une vue d'ensemble de la pile de cartes.

Il publie en **1956** « *A Study of Thinking* », qui tranche avec les principes béhavioristes. Au lieu de s'intéresser aux seuls comportements observables des sujets, il cherche à mettre en évidence les stratégies mentales de sujets confrontés à une tâche.

Cette idée de « **stratégies mentales** » s'écarte radicalement de la perspective béhavioriste encore dominante parce qu'elle s'intéresse aux cheminements de la pensée consciente du sujet qui cherche à résoudre un problème.



Toujours en 1956, **George Miller** publie pour sa part un article au titre énigmatique de « Sept, le nombre magique plus ou moins 2 » qui tentait d'attirer l'attention sur **certaines limites du psychisme humain**.

Par exemple, celle qui fait que l'esprit s'embrouille dès que le nombre d'information à traiter atteint sept (plus ou moins 2).

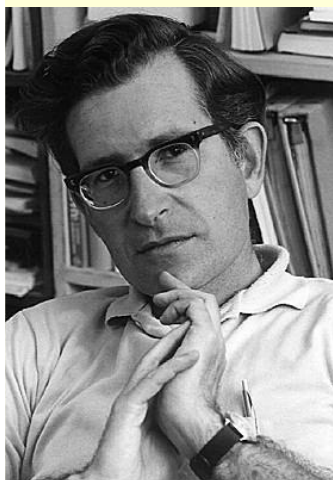
S'il en est ainsi, soutient Miller, c'est que le cerveau possède une **structure propre, avec ses limites** et qu'il ne peut pas être comparé à un réceptacle vierge comme le suppose le béhaviorisme.

De plus, remarque Miller, pour surmonter leurs limites intellectuelles, les humains ont inventé une solution ingénieuse : regrouper les chiffres par grappes.

Miller met ainsi l'accent sur une capacité de l'esprit : celle **d'effectuer un « traitement » logique** qui ne se réduit pas à un simple enregistrement des données transmises.

Ces travaux étaient, à l'époque, **révolutionnaires**, voire **hérétiques**. Ils mettaient l'accent sur les « états mentaux » du sujet, sur ses capacités de raisonnement, sur la façon dont il « traitait l'information ».

Miller sera aussi présent à un autre événement important de l'année 1956 : **un séminaire tenu au MIT** aux côtés de Herbert Simon, Howard Gardner et du jeune Noam Chomsky.



En **1956**, **Noam Chomsky** a 28 ans. Il vient juste de passer sa thèse (*The Logical Structure of Linguistic Theory*) et il est en train de formuler une nouvelle théorie linguistique : **la grammaire générative (GG)** qui est exposée en **1957** dans son ouvrage *Syntactic Structures*.

Chomsky part de l'existence d'une aptitude proprement humaine à produire le langage et à construire des phrases, en associant des mots **selon des règles de grammaire**, règles qui sont universelles, c'est-à-dire communes à toutes les langues.

Ainsi un enfant qui apprend à parler forme des phrases telles que « *tombé ballon* », ou « *ballon tombé* ». Cela signifie qu'il construit des phrases, à partir de constituants fondamentaux (verbe, nom...), organisées selon une règle de production ($P = N + V$).

Par la suite, il enrichira la phrase de nouveaux éléments pour dire « *le ballon est tombé* » ou « *il est tombé, le ballon* ». L'enfant commet des erreurs, mais ces erreurs ont un sens. Il pourra dire « *le ballon a tombé* » ou « *le ballon, il est tombé* ». Mais il applique une règle (l'usage d'un auxiliaire).

Jamais on ne l'entendra dire « *le tombé a ballon* »...

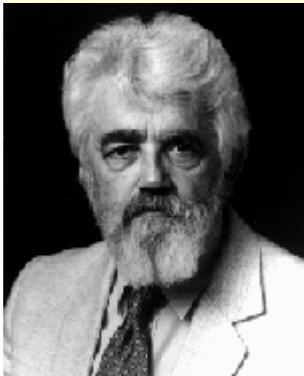
Il semble donc bien y avoir, à la source de la production d'une phrase, une **sorte de logique invisible, des « règles de production » sous-jacentes.**

Ces règles sont sans doute universelles, affirme Chomsky, bien que chaque langue se constitue ensuite à sa manière, en fonction de variations autour de quelques motifs.

Durant **l'été 1956**, au Dartmouth College à Hanover (New Hampshire), se tient **un autre colloque**, le « Dartmouth Summer Research Conference on Artificial Intelligence » considéré comme un événement fondateur pour **l'Intelligence Artificielle** en tant que domaine d'étude.

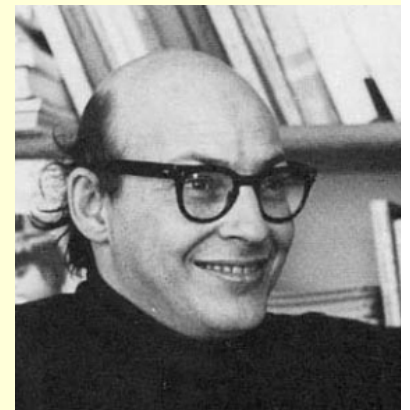
En fait, il s'agissait plutôt, un peu comme les conférences Macy, d'une séance de brainstorming qui a duré un mois complet !

Le projet est fascinant : la mise en oeuvre d'une « intelligence artificielle » capable de **copier** et de **simuler** les performances de **l'intelligence humaine**.



John McCarthy

John McCarthy (qui est décédé le 24 octobre dernier à l'âge de 84 ans), **Marvin Minsky** et Claude E. Shannon, le père de la théorie de l'information, sont présents. Il y a aussi Herbert A. Simon, spécialiste des organisations, et son ami mathématicien Alan Newell.



Ces deux derniers, Simon et Newell, créent la surprise en présentant le premier programme d'intelligence artificielle, destiné à démontrer des théorèmes mathématiques.

Il fonctionne comme une machine logique capable d'articuler entre elles une foule de propositions à partir de quelques prémisses (sur le modèle du syllogisme « si A implique B » et « B implique C » alors « A implique C »).

L'I.A. est née. Son but : **copier, puis dépasser** les activités humaines réputées intelligentes, **comme raisonner, utiliser le langage ou résoudre des problèmes.**

Comment y parvenir ? Simon propose une voie générale : chaque problème à résoudre peut être **décomposé en une série de buts intermédiaires**, et on explorera différentes voies pour chacun d'eux jusqu'à ce que la solution soit trouvée.

Par exemple, pour réparer une panne de voiture, on décompose en deux séries de problèmes : s'agit-il d'un problème électrique ou mécanique ?

Si c'est un problème électrique : vérifier d'abord la batterie, puis le démarreur. Si c'est un problème mécanique, alors etc, etc.

Sur ce principe de résolution de problèmes, Simon et Newell conçoivent, en **1957**, General Problem Solver (GPS), un programme informatique dont la vocation est de résoudre toute une classe de problèmes de même type.

En **1958**, J. McCarthy crée le Lisp (pour *List Processing*, traitement de ligne), un langage de programmation encore très utilisé aujourd'hui.

En **1959**, Arthur Samuel invente un programme de jeu de dames.

Durant les années **1960**, on assiste à l'essor des « systèmes experts ».

En **1965**, Edward Feigenbaum crée *Dental*, un système expert capable de déterminer la formule chimique d'une molécule.

En **1966** naît *Eliza*, un programme de dialogue homme-machine.

En **1970**, paraît le premier numéro de la revue *Artificial Intelligence*.

Simon pense pouvoir bientôt créer une machine à traduire les langues, jouer aux échecs, prendre des décisions, etc. Il ne cache pas son ambition :

« D'ici dix ans, un ordinateur numérique sera champion du monde d'échecs. »

*[ça a pris environ 30 ans de plus, mais c'est arrivé !
En 1997 quand Deep Blue a battu Kasparov]*



Le GSP est vu comme un bon modèle pour rendre compte de la pensée humaine.

*« D'ici dix ans, en psychologie, la plupart des théories prendront la forme de **programmes informatiques** », affirment péremptoirement H.A. Simon et A. Newell.*

Donc l'IA a le vent dans les voiles...



En **1975**, le philosophe **Jerry Fodor** publie ***Le Langage de la pensée*** qui va faire date dans l'histoire des sciences cognitives.

Dans ce livre, Fodor présente un modèle de la pensée qui s'inspire de l'analogie avec le fonctionnement de l'ordinateur. La pensée est au cerveau ce que **le logiciel informatique (*software*) est à la machine (*hardware*)**.

Un programme informatique, comme un calcul mathématique, se présente comme **une série d'instructions**, traitées sous forme d'une **combinaison de symboles liés entre eux par des règles logiques**.

Par exemple, les idées « si les nuages s'accumulent, alors il va pleuvoir » ou « les gens qui ne s'aiment plus devraient divorcer », pourraient être, au fond, réduites à des propositions abstraites du type « si p alors q ».

Tel serait le principe du « **langage de la pensée** » de Fodor : une sorte de langage algébrique, qui convertirait les représentations ordinaires (la pluie, le mariage) en un langage abstrait, fait de symboles (ou « représentations »), sur lesquels s'effectuent des calculs (ou « computations »).

Premier paradigme : Le cognitivism

Les coups portés au behaviorisme dans les années **1950** et le développement des ordinateurs et de l'IA dans les années **1960** permettent l'avènement de l'approche **cognitiviste** (ou **computationnelle**) de l'esprit.

L'ordinateur devient la **métaphore** centrale pour décrire l'esprit humain.
[métaphore dont on connaît aujourd'hui les limites et les méfaits!]

Cette métaphore amène ainsi les cognitivistes à considérer que les éléments du monde extérieur sont traduits en **représentations internes**, exactement comme le font les ordinateurs.

Ces représentations symboliques internes sont ensuite **manipulées selon certaines règles** prédéterminées pour fournir des «sorties» ou des réponses appropriées. Autrement dit, la pensée est un processus de **traitement de l'information**.



Cognitivism (ou symbolisme, ou approche computationnelle) en 3 questions :

[Tiré de Invitation aux sciences cognitives de Varela]

1) Qu'est-ce que la cognition ?

C'est le traitement de l'information : la manipulation de symboles à partir de règles.

2) Comment cela fonctionne-t-il ?

Par n'importe quel dispositif pouvant représenter et manipuler des éléments physiques discontinus : des symboles. Le système n'interagit qu'avec la forme des symboles (leurs attributs physiques), et non leur sens.

3) Comment savoir qu'un système cognitif fonctionne de manière appropriée ?

Quand les symboles représentent adéquatement quelque aspect du monde réel, et que le traitement de l'information aboutit à une solution efficace du problème soumis au système.

Fodor adopte aussi **la thèse dite « fonctionnaliste »**, qu'il reprend de son ancien professeur, le philosophe Hilary Putnam.

*Fodor affirme : « [...] de même qu'un calcul mathématique peut être effectué sur des supports très différents (boulier, machine mécanique ou électronique), on peut **dissocier** l'analyse des opérations effectuées des supports matériels (cerveau ou machine) qui le permettent. »*

L'esprit est organisé en modules spécialisés (comme on va le voir à l'instant) qui peuvent être mis en place sur **d'autres supports** comme les **ordinateurs**.

C'est le fameux concept « de la réalisation multiple ».

Le corps a donc ici une importance marginale. N'importe quelle structure peut faire l'affaire, en autant qu'elle peut effectuer la même fonction, la même "computation" sur des symboles.

Dans sa version computationnelle extrême, le fonctionnalisme soutient que l'incarnation de l'organisme n'a carrément aucune pertinence quant à la nature de son esprit. C'est le **software**, et non le **hardware**, qui compte pour ce qui est de l'activité mentale... [à l'opposé des approches incarnées de la cognition]

Le cognitivisme se raffine...

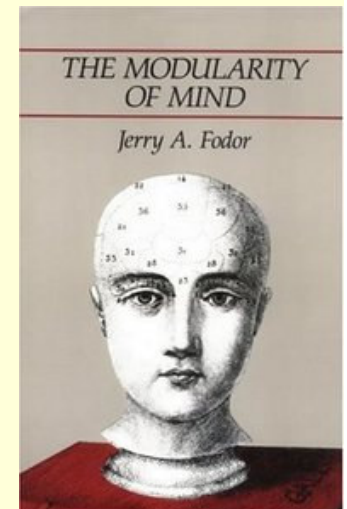
En **1986**, J. Fodor publie ***La Modularité de l'esprit***. Sa thèse centrale remet au goût du jour l'ancienne idée des facultés mentales : l'esprit ne fonctionne pas comme un tout unifié.

Le psychisme humain traite les informations sous forme de **modules spécialisés destinés chacun à un type particulier d'opération**.
« Ainsi, note J. Fodor, les mécanismes perceptifs sont distincts de ceux du langage ou de la mémoire. La maîtrise du langage se décompose à son tour en sous-modules spécialisés : certains sont spécifiques à la grammaire, d'autres traitent des informations sémantiques contenant le sens des mots. »

Pour Fodor, un **module** est donc :

- 1) spécifique à une opération précise ;
- 2) son fonctionnement est autonome, rapide et donc **inconscient** ;
- 3) il possède une localisation neuronale précise.

Mais comment coordonner ces modules entre eux ? Selon Fodor, ils sont sous la coupe d'un **système central** chargé de coordonner et de centraliser les informations traitées par les modules spécifiques.



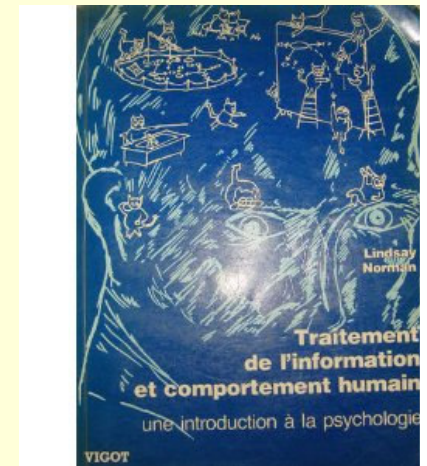
Ce paradigme central du cognitivism allait dominer complètement les sciences cognitives pendant près de **30 ans.** (i.e. de mi-1950 à mi-1980).

Et presque toute la **neurobiologie** (avec son énorme corpus de données empiriques) va être envahie par la perspective du **traitement de l'information** qui caractérise le cognitivisme. La plupart du temps, les origines et les présupposés de cette perspective ne sont même pas questionnés. [c'est évidemment Varela qui parle ici...]

[Exemple: « Traitement de l'information et comportement humain », édition originale anglaise 1977, traduction française 1980, dans lequel j'ai étudié...]

Et l'idée de base selon laquelle **le cerveau est une machine de traitement de l'information** qui répond aux traits de l'environnement reste le **noyau dur des neurosciences** modernes et de leur vulgarisation auprès du public...

[Et je suis certain que vous pouvez trouver abondamment d'exemple de ça dans Le cerveau à tous les niveaux parce que c'est profondément enfoui dans la discipline!]



L'assurance tranquille du paradigme dominant... ;-)

Durant l'âge d'or du cognitivisme dans les années 1970, les cognitivistes aimaient à dire que leur approche était "the only game in town" (Fodor 1975, 1981).

Et ils affirmaient que le modèle de l'ordinateur pour l'esprit n'était pas qu'une simple métaphore mais bien une théorie scientifique ! (Pylyshyn 1984), contrairement à d'autres modèles mécanistes, comme celui du standard téléphonique ("telephone switchboard") l'avait été avant lui...



Et même : **l'argument du « de-quoi-d'autre-pourrait-il-s'agir ? »**
[définition très parlante d'un paradigme dominant...]

Suivant les mots d'Allen Newell :

« Bien qu'il existe une petite chance que nous voyions émerger un nouveau paradigme pour l'esprit, cela me paraît peu probable.
Au fond, il ne semble exister aucune option viable.

Cette position n'est pas surprenante. On aboutit dans bien des sciences à un stade où il n'y a plus guère d'options de rechange aux théories acceptées. Dans ce cas, toute l'activité scientifique intéressante s'inscrit dans la perspective principale.

Il me semble que nous nous rapprochons passablement de cette situation en ce qui a trait à la **théorie computationnelle de l'esprit**. »

Mais...

Mais ! Critiques, problèmes, lacunes, etc... du cognitivisme

*“Human intuition and insight... cannot be reduced to any set of computational rules.” Roger Penrose, *Shadows of the Mind* (1994)*

*“Computation is just interpretable symbol manipulation ; cognition isn’t“
Étienne Harnad (1994)*

Daniel C. Dennett ou **Francisco Varela**, parleront à son propos de la « **grande orthodoxie** » du cognitivisme. (Dennett parlera aussi de « computationnalisme intégriste » !). Et **Varela** reprochera aussi à cette « grande chapelle orthodoxe » de ramener la pensée humaine à un dispositif mécanique et logique alors que **la pensée s'inscrit dans le vivant**.

En particulier, dans la formulation cognitiviste de la « modularité de l'esprit » de Fodor, les processus cognitifs sont considérés comme comprenant des sous-systèmes distincts qui ne peuvent être « pénétrés » par l'activité mentale consciente. Donc qui ne sont pas accessibles au « niveau personnel » subjectif de la **conscience de soi**.

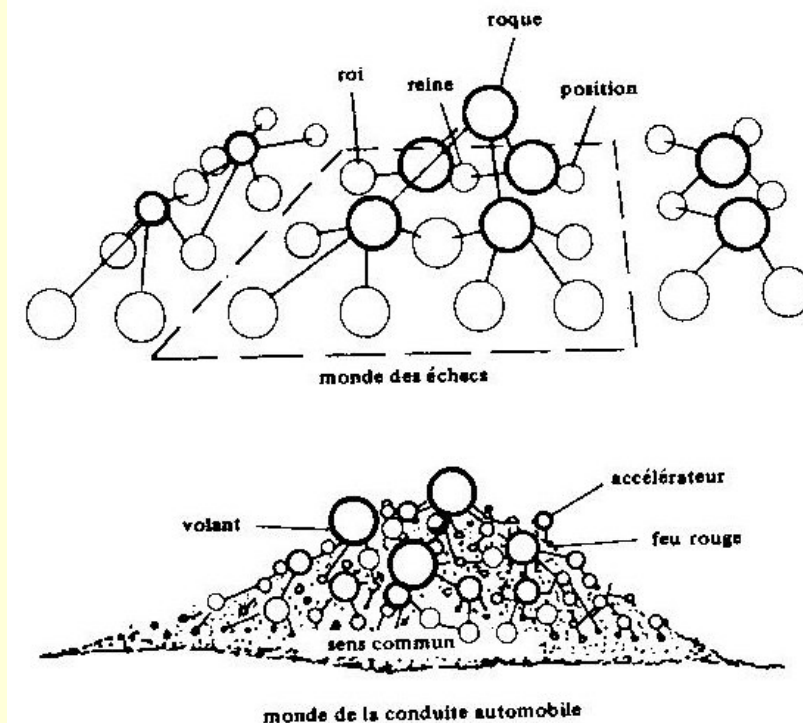
Donc on retrouve une espèce de **tabou de la conscience** qui n'est pas sans rappeler un certain behaviorisme... Ce qui est inacceptable pour certains.

Autre problème, celui du cadre de référence

Varela : **la plus simple action cognitive** exige un nombre de connaissances apparemment **infini** que nous prenons pour acquises, mais qui doivent être introduites dans l'ordinateur en détail, une par une.
[= « **frame problem** », cf. le schéma ci-bas]

Il a donc fallu **abandonner l'espoir**, nourri par les premiers cognitivistes, **de réaliser une « machine à résoudre les problèmes généraux »** (GPS, general problem solver).

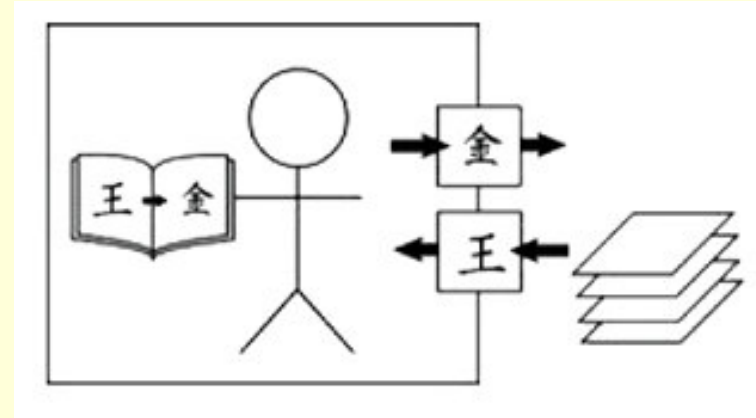
On a dû se contenter de programmes fonctionnant dans des domaines de connaissance locaux, où le programmeur peut introduire dans la machine la quantité nécessaire de ces **connaissances d'arrière-plan**.



Autre problème, celui du sens

A partir des **années 1980**, le philosophe **John R. Searle**, développe une série d'arguments pour démontrer que la machine **ne pense pas** car elle **n'a pas accès au sens**.

L'argument de la « chambre chinoise » est destiné à montrer qu'une machine ne fait que manipuler des symboles abstraits, sans en comprendre la signification.



Elle peut traduire mot à mot un texte dans deux langues étrangères, dès lors qu'elle dispose d'un dictionnaire de correspondances.

Mais ne comprenant pas le sens des mots utilisés, elle bute donc sur les ambiguïtés sémantiques : comment peut-on choisir entre « *weather* » ou « *time* » pour traduire le mot français « temps », si on n'a pas accès à son sens ?

La critique est forte, car c'est justement dans le domaine de la traduction automatique que l'IA, dont s'inspire le cognitivisme, montre ses limites...

Deux autres déficiences largement reconnues du cognitivisme

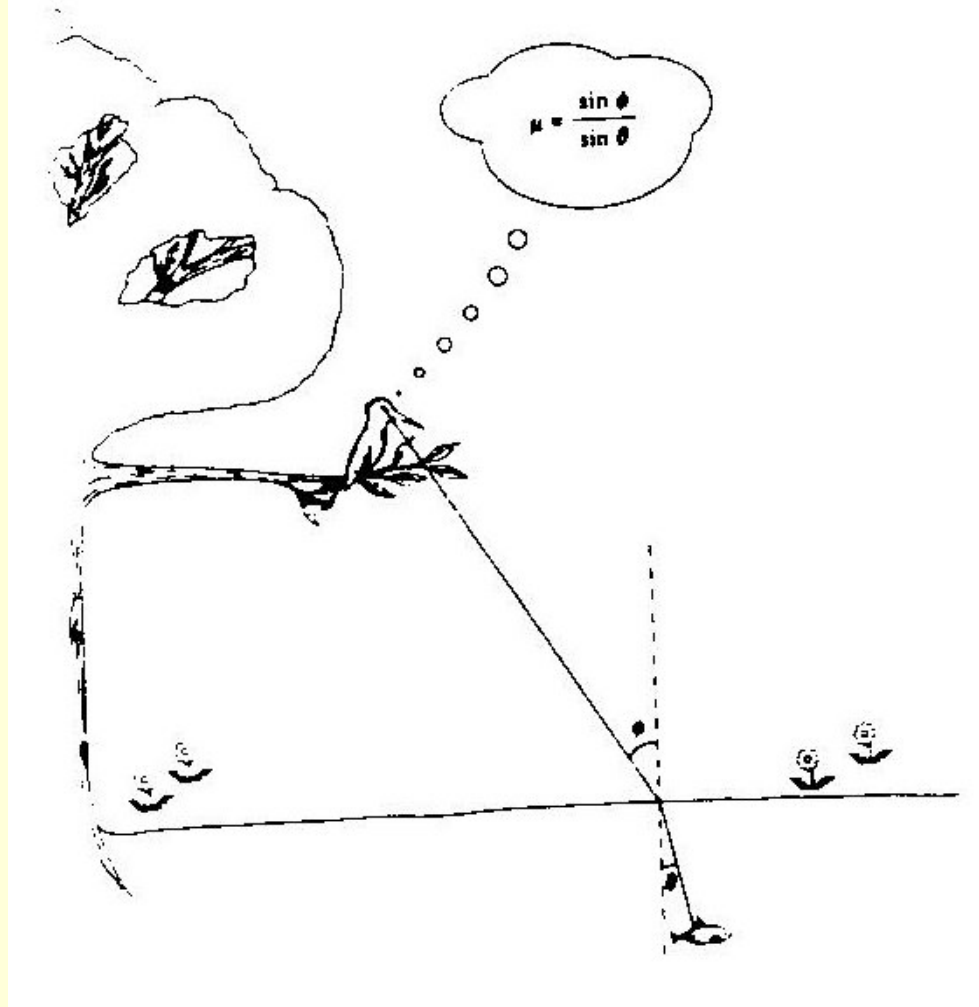
La première est que le traitement de l'information symbolique repose sur des règles **séquentielles**, appliquées une par une.

Cela crée un « **goulot d'étranglement** » qui constitue une grave limitation quand la tâche à accomplir requiert un grands nombres d'opérations séquentielles.

(comme l'analyse d'image ou la prévision météorologique)

Une seconde limitation importante réside dans le fait que le traitement symbolique est **localisé** : la perte ou le dysfonctionnement d'une partie quelconque des symboles ou des règles du système provoque un sérieux dysfonctionnement de l'ensemble.

À l'inverse, un mode d'opération **distribué** est hautement souhaitable, car il garantit au moins une équipotentialité et une immunité relative face aux mutilations...



Enfin : pour capturer sa proie, du point de vue cognitiviste, un martin pêcheur doit avoir dans son cerveau la représentation de la loi de la réfraction de Snell...

[Tiré de Invitation aux sciences cognitives de Varela]

Vers le connexionnisme...

Le cognitivisme voulait atteindre des performances comparables à celles d'un expert humain hautement formé. Mais comme il ne réussissait bien qu'à résoudre de tâches plus circonscrites et locales, une conviction s'est développé : la forme **d'intelligence** la plus fondamentale n'est pas celle de l'expert, mais bien celle d'un **bébé** !

Car un bébé peut acquérir le langage à partir d'énonciations quotidiennes dispersées et constituer des objets signifiants à partir de ce qui semble être une masse informe de stimuli.

Il fallait donc intervertir les deux dans nos démarches et chercher plutôt à simuler l'intelligence du bébé qui apprend.

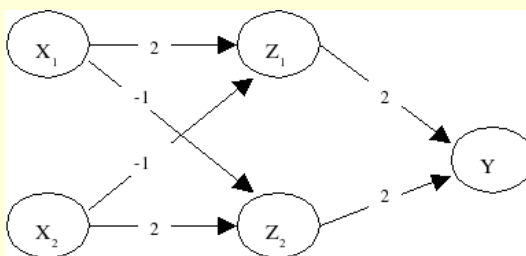
On en vint aussi à considérer que les architectures cognitivistes s'étaient trop éloignées des intuitions biologiques : les tâches les plus ordinaires, y compris celles qu'accomplissent les insectes (ou les martin-pêcheurs!), sont effectuées avec une rapidité que l'on ne peut pas atteindre en s'appuyant sur le type de stratégie computationnelle que propose l'orthodoxie cognitive.

Même chose pour la capacité du cerveau biologique à maintenir sa fonctionnalité à la suite de dommages importants.

Ou encore la souplesse de la cognition biologique qui permet de s'adapter à de nouveaux environnements sans perdre toutes ses compétences (autre situation où le paradigme computationnel ne réussit pas bien...)

Bref : **le cerveau va redevenir la source principale des métaphores.**

Et à **la fin des années '70**, on va redécouvrir les idées autour des « réseaux de neurones », déjà mises de l'avant par Mc Culloch autour des années 1950.



À n'en pas douter, l'un des facteurs qui a contribué à ce regain d'intérêt est la **redécouverte parallèle des idées d'auto-organisation** en physique et en mathématiques non linéaires, ainsi que **la facilité d'accès aux ordinateurs rapides.**

Les apports de cette nouvelle métaphore du « réseau de neurones »

Déjà aux conférences Macy on étudia de manière approfondie le fait que, dans les cerveau concrets, il semble n'y avoir

ni règles,

ni dispositif logique central de traitement.

Et que l'information ne paraît pas engrangée à des adresses précises.

Au contraire, on peut considérer que les **cerveaux** opèrent de manière **distribuée** sur la base **d'inter-connexions massives**.

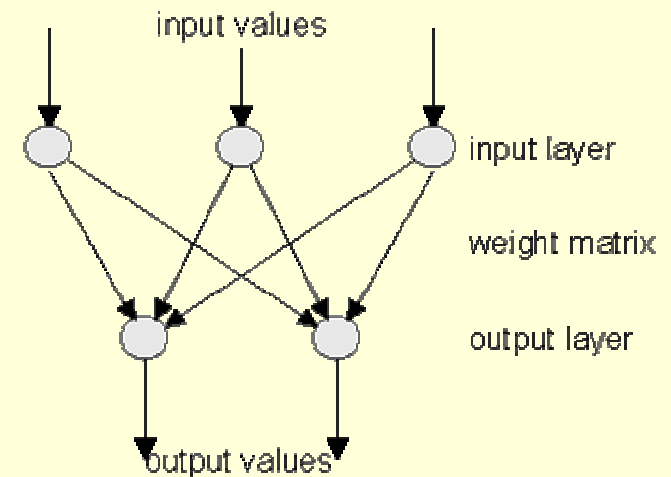
De sorte que les connexions effectives entre les ensembles de neurones **se modifient** en fonction du déroulement de l'expérience.

Bref, ces ensembles présentent une **capacité auto-organisatrice** dont rien ne peut rendre compte dans le paradigme la manipulation de symbole.

Deuxième paradigme : Le connexionnisme

L'expression « **connexionnisme** » fut déjà utilisée par **Rosenblatt en 1958** (plus précisément, c'était l'adjectif « connectionist »).

En **1958**, Frank Rosenblatt construit le « **Perceptron** », appareil très simple doué de certaines capacités de reconnaissance, sur la seule base de **changement de connectivité** entre des composants analogues à des neurones.

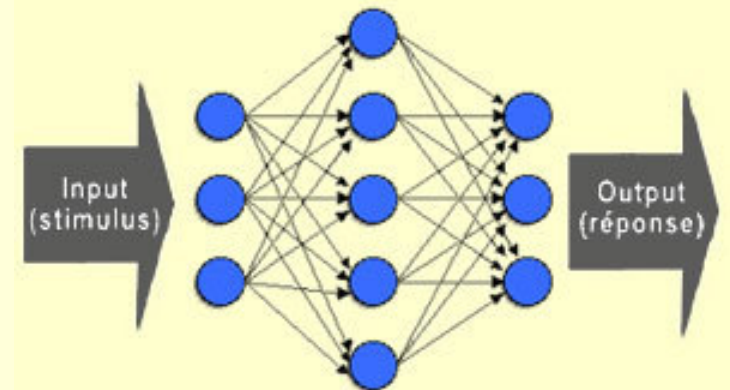


Il ne s'agit donc pas d'une **modélisation logique (top down)**, mais bien de l'exploration des ressources d'un **réseau complexe d'éléments simples en interaction (bottom up)**.

L'analogie principale du connexionnisme est donc le **réseau de nombreuses unités** (ou « neurones artificiels ») **interconnectées entre elles**.

Cette approche, basée sur le **traitement en parallèle** de l'information, s'est donc développée avec le souci de **rapprocher la structure des modèles cognitifs de celle du cerveau**.

Il s'agit donc d'une approche du **bas vers le haut**, où l'on cesse de parler de représentations en terme de symboles pour les analyser plutôt en termes de **liens** entre de nombreux agents **distribués, coopératifs et auto-organisés**.



Marvin Minsky, qui a inspiré cette approche, considère ainsi le système cognitif comme une société de **micro-agents** susceptibles de résoudre des problèmes localement.

Pour les connexionnistes, il faut donc aller en deçà des opérations symboliques, vers ce qu'ils appellent le niveau « **subsymbolique** ».

[à ne pas confondre avec le niveau « subpersonnel », i.e. inconscient (mais symbolique) du cognitivisme...]

Connexionnisme en 3 questions :

[Tiré de Invitation aux sciences cognitives de Varela]

1) Qu'est-ce que la cognition ?

C'est l'émergence d'états globaux dans un réseau de composants simples.

2) Comment cela fonctionne-t-il ?

Des règles locales gèrent les opérations individuelles et des règles de changement gèrent les liens entre les éléments.

3) Comment savoir qu'un système cognitif fonctionne de manière appropriée ?

Quand les propriétés émergentes (et la structure résultante) sont identifiables à une faculté cognitive – une solution adéquate pour une tâche donnée.

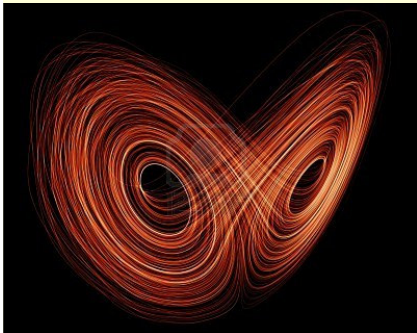
Sur l'émergence ...

Pour les connexionnistes, les **représentations symboliques** n'apparaissent qu'au niveau **collectif**, i.e. elles « **émergent** » des interactions entre les nombreuses unités de computations simples.

(et la **représentation** se ramène donc à une correspondance entre cet état global émergent et des propriétés du monde)

Ce passage de règles locales à une cohérence globale est au cœur de ce que l'on avait coutume d'appeler « **auto-organisation** » pendant les années cybernétique. Mais aujourd'hui, on préfère parler de **propriétés émergentes** ou globales, de dynamique des réseaux, de réseaux non linéaires ou de systèmes complexes.

Des propriétés émergentes ont été découvertes dans tous les domaines : oscillations chimiques, génétique des population, réseaux immunitaires, écologie, géophysique, etc.



Et l'une des manières les plus fécondes de capter les propriétés émergentes communes à ces divers systèmes consiste à utiliser **la notions d' « attracteur »** élaborée dans la théorie des systèmes dynamiques.



Rappel d'autres différences avec le cognitivisme...

Il n'y a plus ici d'algorithmes complexes menés séquentiellement par différents modules comme dans le cognitivisme.

Pour les connexionnistes, la conception en série du traitement de l'information n'est pas compatible avec la rapidité avec laquelle nous sommes capables de traiter une information.

Et vu la lenteur de l'influx nerveux, il faut postuler un traitement parallèle pour expliquer la rapidité d'une réaction.

Par exemple, si nous traitons en série les caractéristiques d'un obstacle sur notre route, puis la nécessité ou non de freiner, et la façon dont on enfonce la pédale de frein, il est certain qu'au moment d'agir, il sera trop tard.

De plus, le paradigme connexionniste refuse la notion de modularité.

On n'a donc plus besoin **d'un centre de contrôle ou d'une unité centrale** pour traiter le tout puisque les **réseaux de neurones distribués en parallèle**, comme dans le cerveau, semblent pouvoir très bien s'en passer.

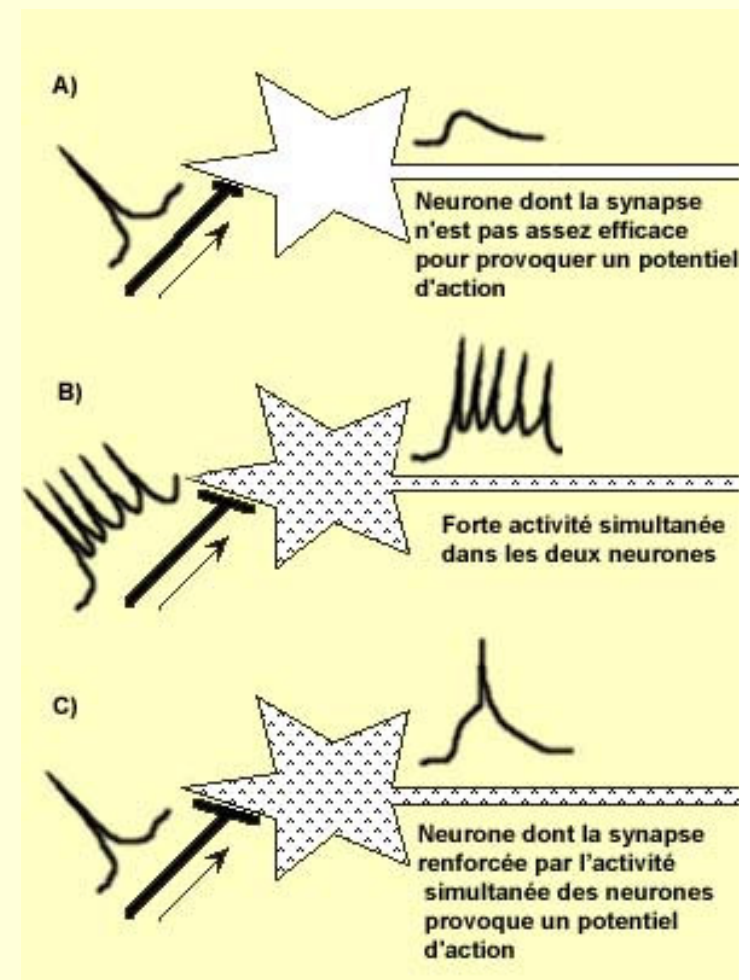
[on note cependant ici que toutes les opérations continuent de se faire exclusivement dans le cerveau, sans référence au corps...]

Autre différence importante : **l'efficacité des connexions** qui unit les éléments du réseau **se modifie en fonction de l'expérience.**

Cela s'obtient d'habitude par le biais d'une règle déterminant le changement graduel des connexions à partir d'un état initial relativement arbitraire. La règle d'apprentissage que l'on a étudiée le plus à fond est la « **règle de Hebb** ».

En **1949**, Donald Hebb suggéra que l'apprentissage peut reposer sur des changements qui, à l'intérieur du cerveau, découlent du degré de corrélation de l'activité des neurones : si deux neurones ont tendance à être activés simultanément, leur connexion est renforcée; dans le cas contraire, la force de la connexion est diminuée.

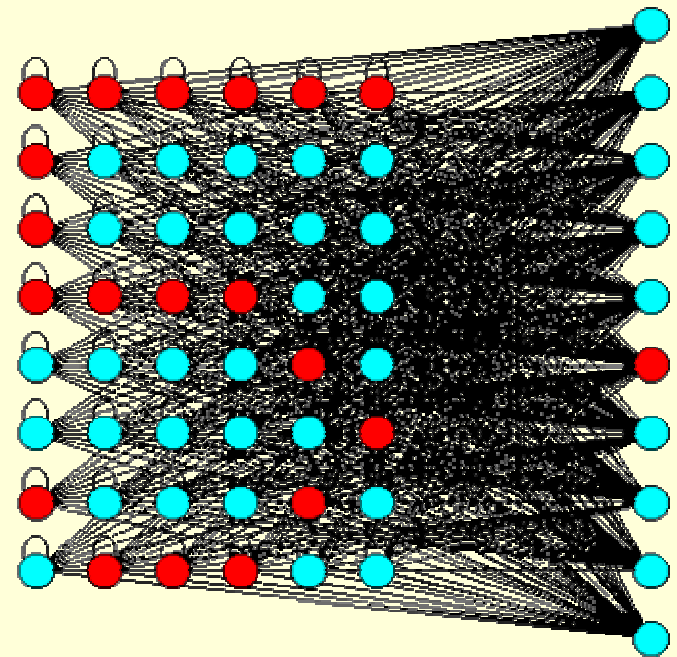
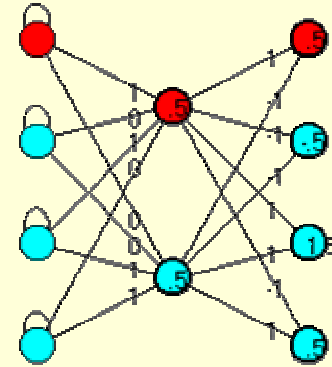
C'est pourquoi la connectivité du système (et donc sa façon de « computer ») devient alors **inséparable de l'histoire de sa transformation**.



Contrairement au cognitivisme ou à l'intelligence artificielle traditionnelle où toutes les opérations devaient être écrites à l'avance par un programmeur, **les réseaux de neurones artificiels ne sont pas programmés, mais plutôt entraînés**.

Et pour bien des tâches, comme pour la reconnaissance des formes, cela s'est avéré une approche fructueuse.

On montre par exemple plusieurs fois la forme à reconnaître à la couche d'entrée du réseau de neurone, et au bout d'un certain nombre d'essais, grâce aux règles locales de modification de l'efficacité synaptique (de type hebbien), le système « ***apprend*** » à ***modifier le « poids » de chacune de ses connexions*** et à fournir une réponse appropriée issue de l'activité ayant émergé dans son réseau.

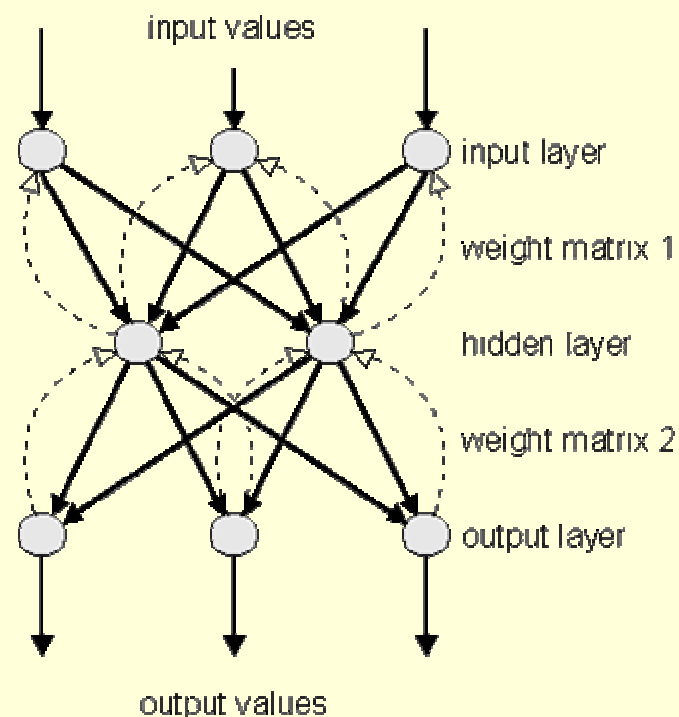


Au début des années 1980, David Rumelhart et David Parker ont indépendamment redécouvert un vieil algorithme : la technique de **rétropropagation du gradient** (*Backpropagation* en anglais), qui permet d'améliorer la performance des réseaux de neurones.

Cette méthode permet de **corriger les erreurs** selon l'importance des neurones virtuels qui ont justement participé à la réalisation de ces erreurs, de la dernière couche vers la première.

Les poids synaptiques qui contribuent à engendrer une erreur importante se verront alors modifiés de manière plus significative que les poids qui ont engendré une erreur marginale.

À chaque essai durant l'entraînement du réseau de neurone, **l'erreur globale tend donc à décroître et la reconnaissance du système à s'améliorer.**



S'améliorer, apprendre, c'est donc pour le système « tomber » pour ainsi dire dans un pattern donné, dans un état global ou une configuration interne unique (un « attracteur ») qui va alors « représenter » l'objet appris.

[Et ces systèmes distribués peuvent réaliser une reconnaissance correcte même si le modèle est présenté partiellement, ou avec du bruit de fond, ou encore si le système lui-même est partiellement mutilé.]

Donc pour le connexionnisme, **le sens** n'est pas contenu dans le symbole, mais il **émerge** de l'activité d'un ensemble d'unités **inférieures au symbole**, appelé **niveau sub-symbolique**.

Le sens n'existe pas non plus dans chaque unité sub-symbolique prise isolément, mais dans la **configuration particulière de ces unités lors de leur activation simultanée**.

Une idée, un concept, ne serait donc que le résultat de l'activation simultanée d'un sous-réseau particulier du système grâce à l'efficacité accrue de ses connexions.

Et ce réseau (et donc ces idées et ces concepts) s'organiserait progressivement, au fur et à mesure de nos expériences.

Le connexionnisme accorde beaucoup d'importance à la **constitution du sens des symboles tout au long de l'histoire du système**, car dans la vie de tous les jours, nous identifions très aisément les objets qui nous entourent et nous apprenons à leur donner un sens.

Donc ici l'intelligence est associée à **la reconnaissance et à l'émergence du sens** (i.e. de catégorie sémantique), alors que pour le cognitivisme c'était plutôt le raisonnement déductif pour résoudre des problèmes.

Et quand un réseau est entraîné et devient efficace pour une tâche, personne ne sait dans les détails comment le réseau trouve les bonnes réponses !

Car contrairement aux logiciels traditionnels, les réseaux connexionnistes ne font pas ce que des programmeurs leur ont dit de faire.

Ils inventent eux-mêmes une partie de la stratégie gagnante.

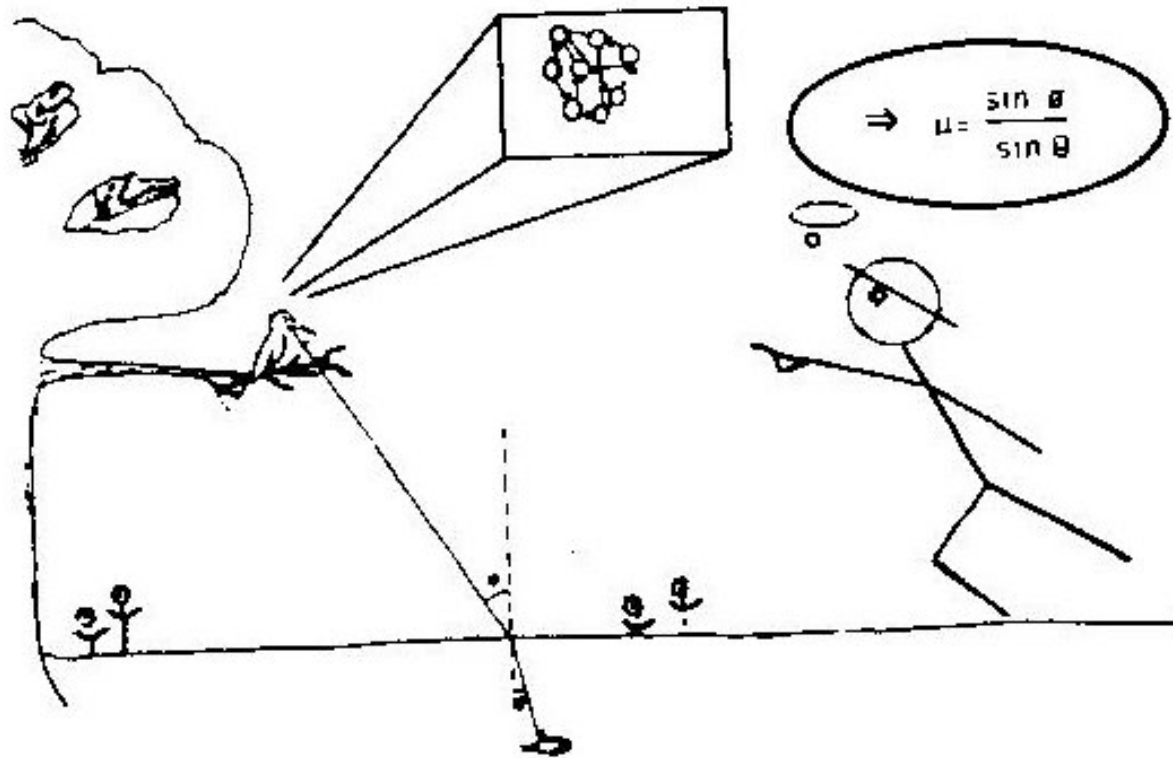


Figure 8 : Dessin de *Punch* : version modifiée. (Voir fig. 3.)

Si on revient à notre martin-pêcheur, **les symboles ne sont plus considérés que** comme des macro-descriptions approximatives d'opérations dont les principes gouvernants appartiennent au niveau sub-symbolique.

En résumé, l'enthousiasme pour le connexionnisme dans les années 1980 était venu entre autres du fait que :

L'IA et les neurosciences cognitives disposaient de peu de résultats convaincants pour expliquer des processus cognitifs comme la reconnaissance rapide, la mémoire associative ou la généralisation catégorielle.

Les modèles connexionnistes étaient beaucoup plus proches des systèmes biologiques...

Les modèles connexionnistes étaient suffisamment généraux pour pouvoir être appliqués, sans guère de modifications, à des domaines variées tels que la vision ou la reconnaissance de la parole.

Mais les modèles connexionnistes vont s'essouffler à leur tour.

Séduisants sur le plan conceptuel, ils tardent à réaliser de véritables simulations des opérations mentales complexes.

De sorte qu'on va proposer toutes sortes d'hypothèses sur la relation possible qu'on pourrait établir entre **cognitivism** et **connexionnisme** pour pallier à leur limites mutuelles (i.e. pas mutuellement exclusifs) :

- la « **ségrégation** » (chacun s'occuperait de « ce dans quoi il est bon »)
- la « **compilation** » : relègue les réseaux de neurones au rôle subalterne où ils « réalisent » au niveau le plus bas les modèles cognitivistes (comme le « langage machine » des ordinateurs par rapport aux commandes données en langage de haut niveau)

Pour Varela et plusieurs autres, la relation la plus intéressante est

- « **l'inclusion** » : considérer les symboles comme une description de niveau supérieur de propriétés qui se trouvent, en dernière instance, enracinées dans un système distribué sous-jacent.

L'explication d'une capacité cognitive comme **la mémoire** intègre depuis quelques temps certaines notions du symbolisme et du connexionnisme. Une conception en termes de modules ne peut pas être complètement écartée, mais il faut en même temps décrire des mécanismes au sein de ces modules sous forme de processus parallèles et distribués.

Mais cela ne règle pas tout...

Plusieurs ne sont pas à l'aise avec **de nombreux presupposés du connexionnisme.**

Les systèmes connexionnistes ne comprennent **aucun couplage sensori-moteur avec l'environnement.**

Au contraire, ils opèrent sur la base d'entrées et de sorties artificielles (encadré par le designer du système).

Le connexionnisme a également hérité du cognitivisme l'idée que la cognition est fondamentalement la **résolution de problèmes prédéfinis** et posés au système **de l'extérieur** par le designer du système ou un observateur.

Le connexionnisme considère que la pensée réside exclusivement dans les **processus inconscients** du domaine subpersonnel.

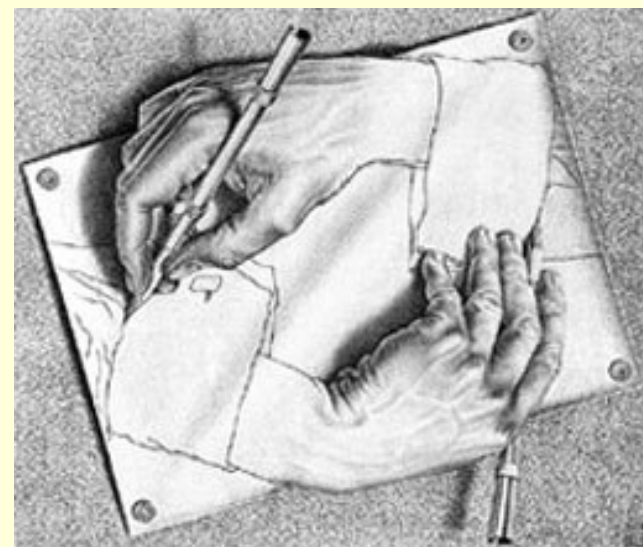
Enfin, il n'apporte que très peu de nouvelles avenues pour réduire le "fossé explicatif" ("explanatory gap", en anglais) entre les computations de la cognition et l'aspect phénoménologique, conscient, de la cognition humaine.

Tant avec l'approche cognitive que avec l'approche connexionniste, la notion de **représentation** allait aussi devenir de plus en plus problématique.

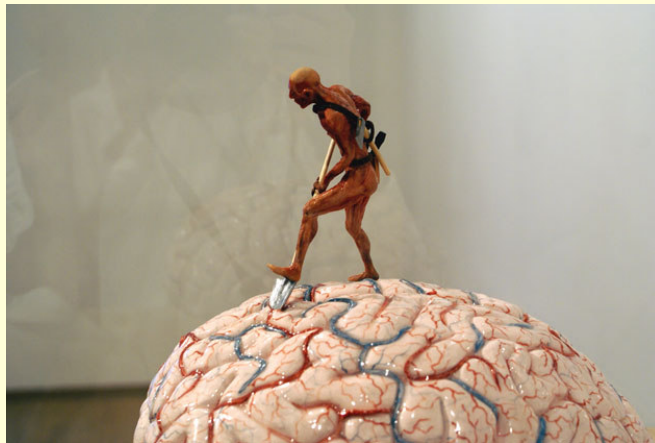
Et en particulier l'idée implicite des représentations qu'elles sont **stables et fidèles** à certains éléments du **monde extérieur**.

Car selon ce que Marvin Minsky lui-même disait à l'époque, et que l'on sait encore plus pertinemment aujourd'hui, c'est que l'activité principale du cerveau consiste en fait à **opérer en permanence des modifications sur lui-même**.

Ce que nous vivons aujourd'hui influencera le rappel d'un souvenir qui, loin d'être toujours le même, sera **une reconstruction** à partir de l'état actuel du cerveau. Et ce souvenir reconstruit affectera inévitablement le fonctionnement subséquent du cerveau.



Par conséquent, contrairement à une machine qui fabrique un objet qui n'a aucun effet sur le fonctionnement de la machine, le cerveau est une machine dont les processus modifient en permanence le fonctionnement subséquent de ladite machine.



Bref avec le cerveau, les résultats des processus deviennent les processus eux-mêmes.

C'est pour cela que plusieurs vont proposer qu'au lieu de représenter un monde indépendant, on peut voir nos processus cognitifs comme faisant plutôt **émerger un monde**, comme quelque chose **d'inséparable des structures dans lequel s'incarne le système cognitif.**

Voilà ce qui a amené certains chercheurs à mettre en doute sérieusement l'existence d'un monde prédonné, duquel le système cognitif devrait **extraire de l'information.**

C'est contre cette métaphore tenace d'un agent cognitif qui ne saurait survivre qu'en possession de la carte d'un monde extérieur que Francisco Varela a élaboré sa théorie de **l'énaction**.

(qui fait partie de la grande famille des systèmes dynamiques)

[et je vous rassure tout de suite, pour les tenants de cette conception des choses, le monde physique existe bel et bien!

Ce n'est pas la réalité physique qu'ils remettent en question, mais bien la façon dont nous sommes contraints de l'appréhender]

(on va prendre tout le prochain cours pour essayer d'expliquer tout cela, en parlant par exemple de la vision des couleurs, etc.)

