

# IA & Cognition Approche cognitiviste

Marie Lefevre

Université Lyon 1 - Laboratoire LIRIS

[marie.lefevre@liris.cnrs.fr](mailto:marie.lefevre@liris.cnrs.fr)



# Cognition



- ▶ Processus par lequel un organisme acquiert la conscience des événements et objets de son environnement
- ▶ Par extension, il réfère à la capacité de raisonner, de percevoir, de décider, de résoudre des problèmes
- ▶ Donc à toute activité que l'on peut associer à *l'intelligence*.

# Processus cognitifs

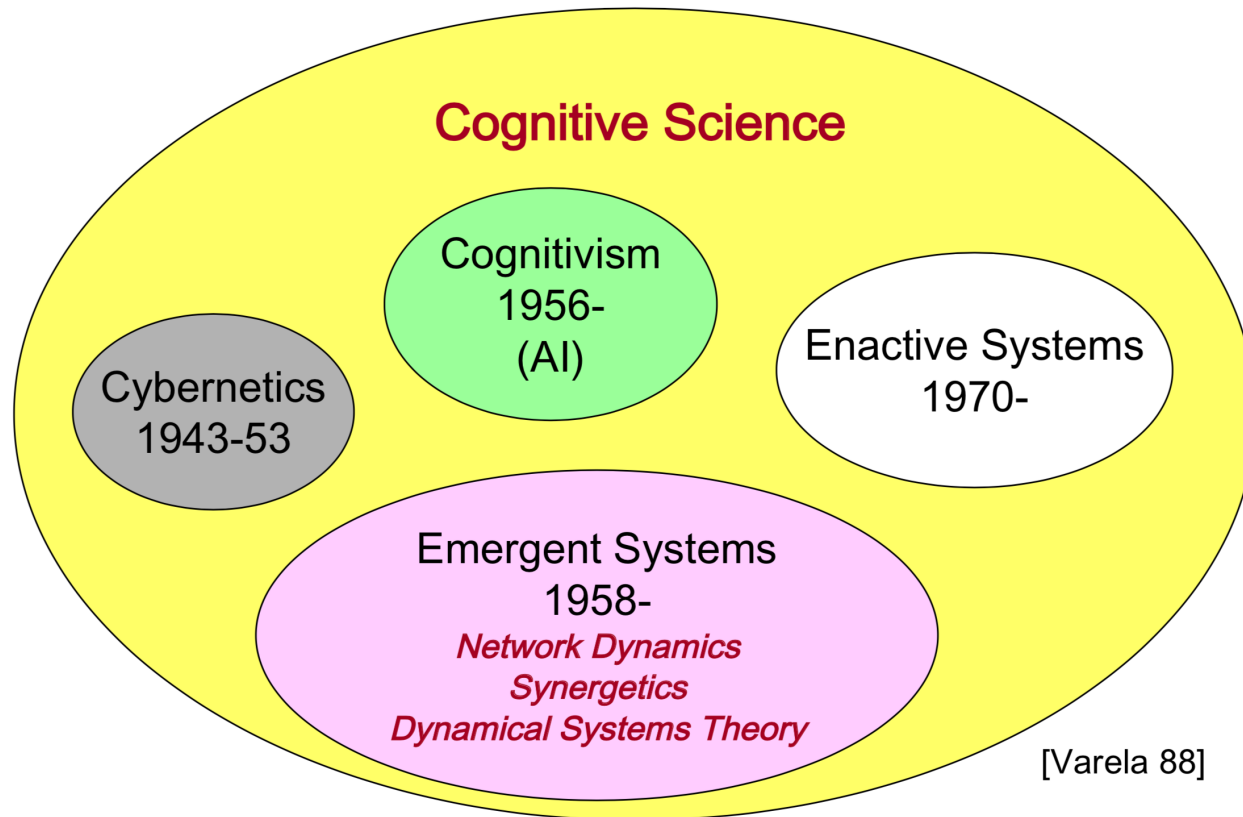
- ▶ Les **processus cognitifs** sont les différents modes à travers lesquels un système traite l'information en y répondant par une action
- ▶ Deux types de systèmes capables de réaliser des processus cognitifs :
  - ▶ les systèmes naturels : un neurone, un réseau de neurones, un cerveau (humain ou animal), un groupe d'individus (poissons, fourmis), etc.
  - ▶ les systèmes artificiels : réseau de neurones artificiels, système expert, etc.
- ▶ Le **traitement de l'information** se définit comme étant le processus par lequel l'information perçue est analysée et intégrée dans la structure de connaissances de la personne / l'agent
- ▶ Il est analysé selon deux dimensions :
  - ▶ le mode de traitement
  - ▶ le niveau d'élaboration



# Processus cognitifs

- ▶ Chez les êtres humains, les processus cognitifs renvoient donc à des enchaînements d'opérations mentales en relation avec la saisie des informations, leur stockage et leur traitement
- ▶ Ces processus s'appliquent :
  - ▶ Perception - Attention - Sensation
  - ▶ Mémoire - Représentation (du monde, réel ou fictif, dans lequel on se place pour raisonner)
  - ▶ Interaction - Langage
  - ▶ Raisonnement - Catégorisation - Prise de décision - Reconnaissance
  - ▶ Apprentissage - Émotion - Oubli
  - ▶ Action - Comportement individuel et collectif
  - ▶ Phénomènes collectifs

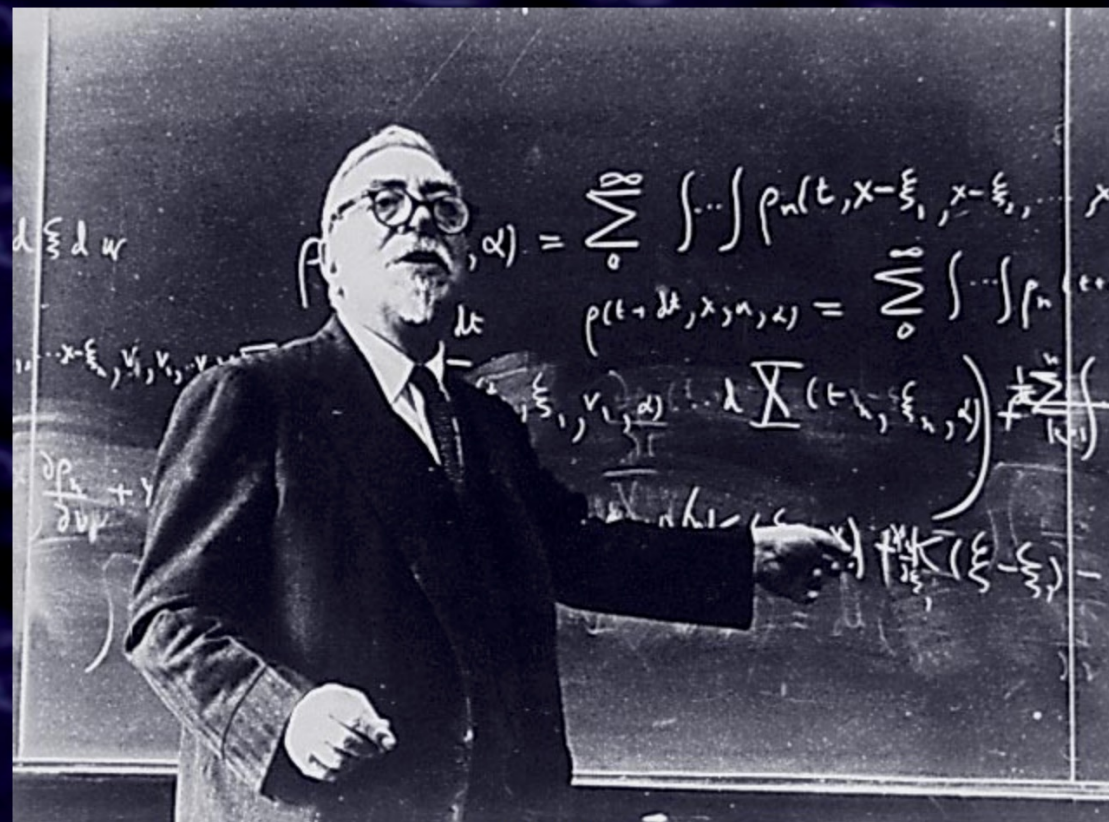
# Paradigmes / courants en Sciences Cognitives



# Qu'est-ce que la Cybernétique ?

- ▶ Conférences Macy, entre 1942 et 1953
- ▶ Groupe interdisciplinaire de mathématiciens, logiciens, anthropologues, psychologues et économistes ...
- ▶ Objectif : édifier une science générale du fonctionnement de l'esprit
  - ▶ Articulé autour du concept clé de rétroaction (feedback)
  - ▶ Donné une vision unifiée de différents domaines ('automatique, électronique, économie...)

## Cybernétique, la science des systèmes



Norbert Wiener

La cybernétique est une science du contrôle des systèmes, vivants ou non-vivants, fondée en 1948 par le mathématicien américain Norbert Wiener.

Notre monde est intégralement constitué de systèmes, vivants ou non-vivants, imbriqués et en interaction. Peuvent ainsi être considérés comme des "systèmes": une société, une économie, un réseau d'ordinateurs, une machine, une entreprise, une cellule, un organisme, un cerveau, un individu, un écosystème...

Les ordinateurs et toutes les machines intelligentes que nous connaissons aujourd'hui sont des applications de la cybernétique. La cybernétique a aussi fourni des méthodes puissantes pour le contrôle de deux systèmes importants: la société et l'économie. (voir le document sur les "Armes silencieuses", un manuel de programmation de la société directement inspiré par la cybernétique)

# Cybernétique

- ▶ Wiener déclare avoir fait dériver le mot *cybernétique* du mot grec *kubernetes* qui signifie *pilote / gouverneur*
- ▶ Le terme est défini dans son livre « Cybernetics » publié en 1948 comme « *the science of control and communication* »
- ▶ W.R. Ashby décrit la cybernétique, dans son livre « An Introduction to Cybernetics » (1956) comme « *the art of steersmanship* »
  - ▶ i.e l'art du pilotage d'une machine par rétroaction
  - ▶ coordination, régulation, contrôle

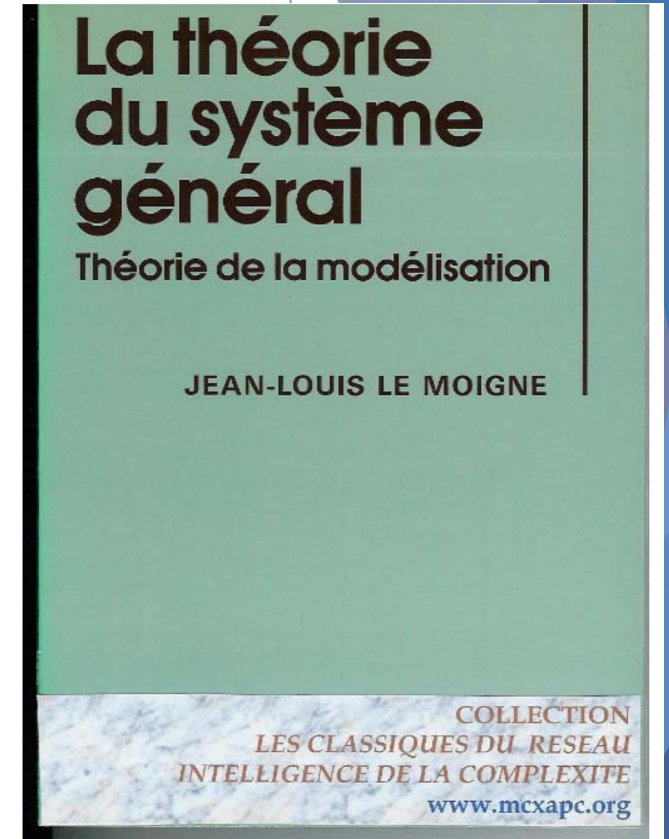
# General System Theory / Systemics

**Ludwig von Bertalanffy**

passages from  
General System Theory

(1968)

<https://www.panarchy.org/vonbertalanffy/systems.1968.html>



<http://www.mcxapc.org/inserts/ouvrages/0609tsgtm.pdf>

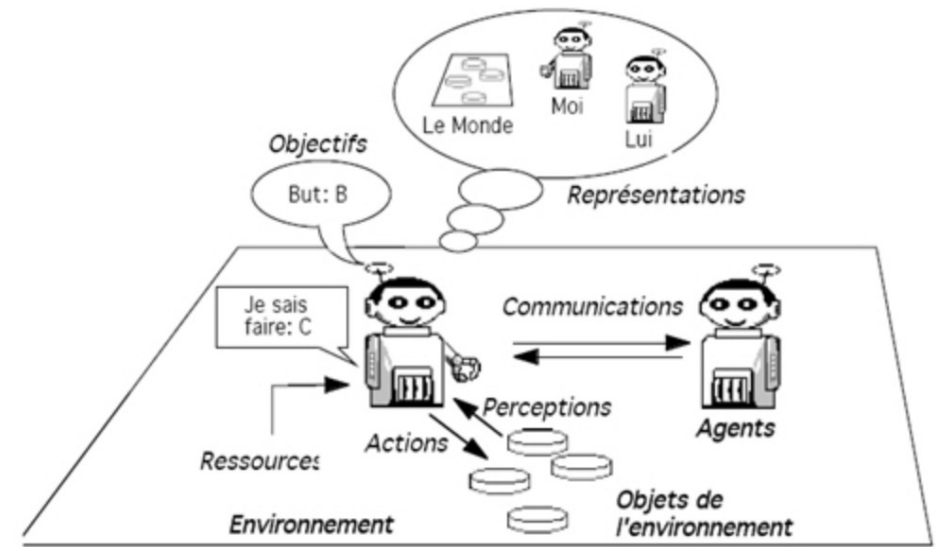


# Closed-loop Control

- ▶ Une action du système provoque un certain changement dans son environnement et ce changement est transmis au système *via* une rétroaction qui permet au système de modifier son comportement
- ▶ Cette relation « causale circulaire » est nécessaire et suffisante pour parler de système dans une perspective cybernétique
- ▶ L'objectif essentiel de la cybernétique est de comprendre et de définir les fonctions et les processus des systèmes qui ont des objectifs et qui participent à des chaînes causales circulaires qui vont de l'action à la détection à la comparaison avec l'objectif souhaité, et à nouveau à l'action

# Cybernétique : 1943-1953

...  
mais les idées perdurent

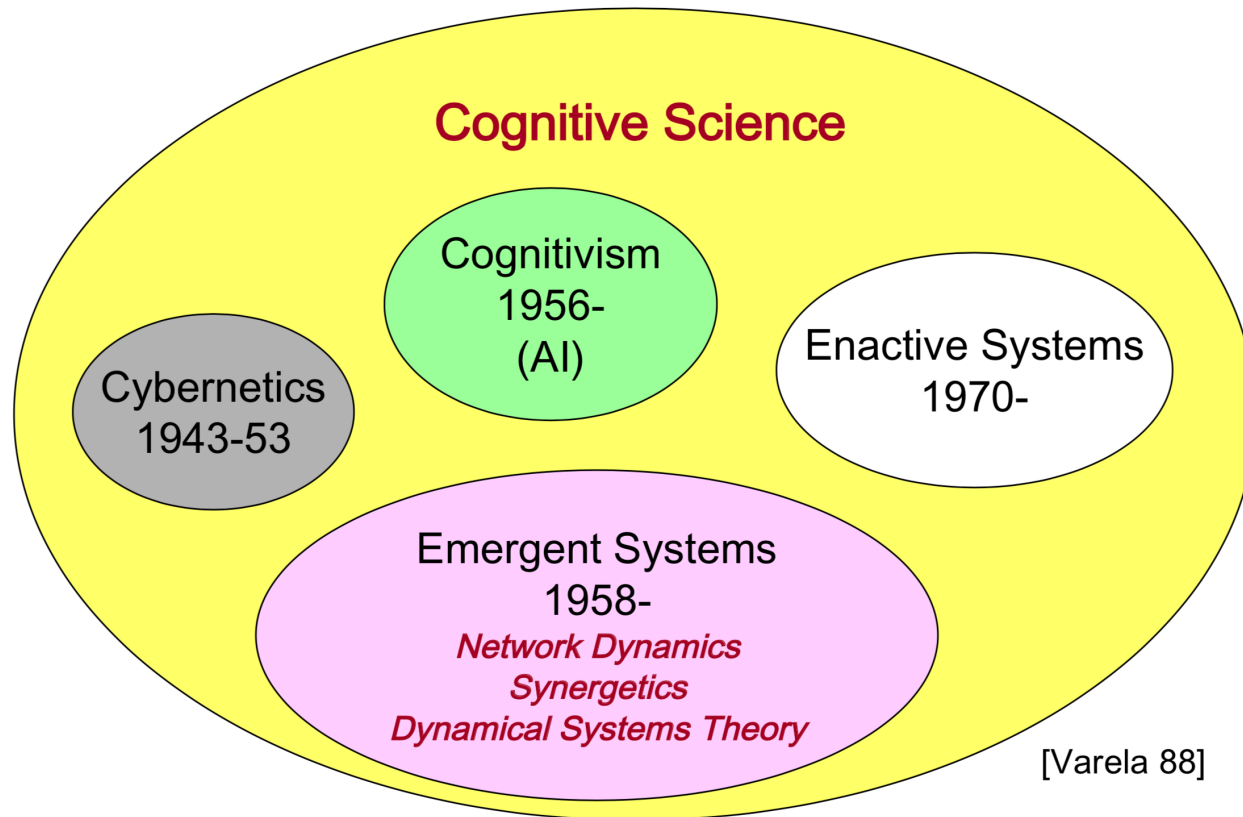


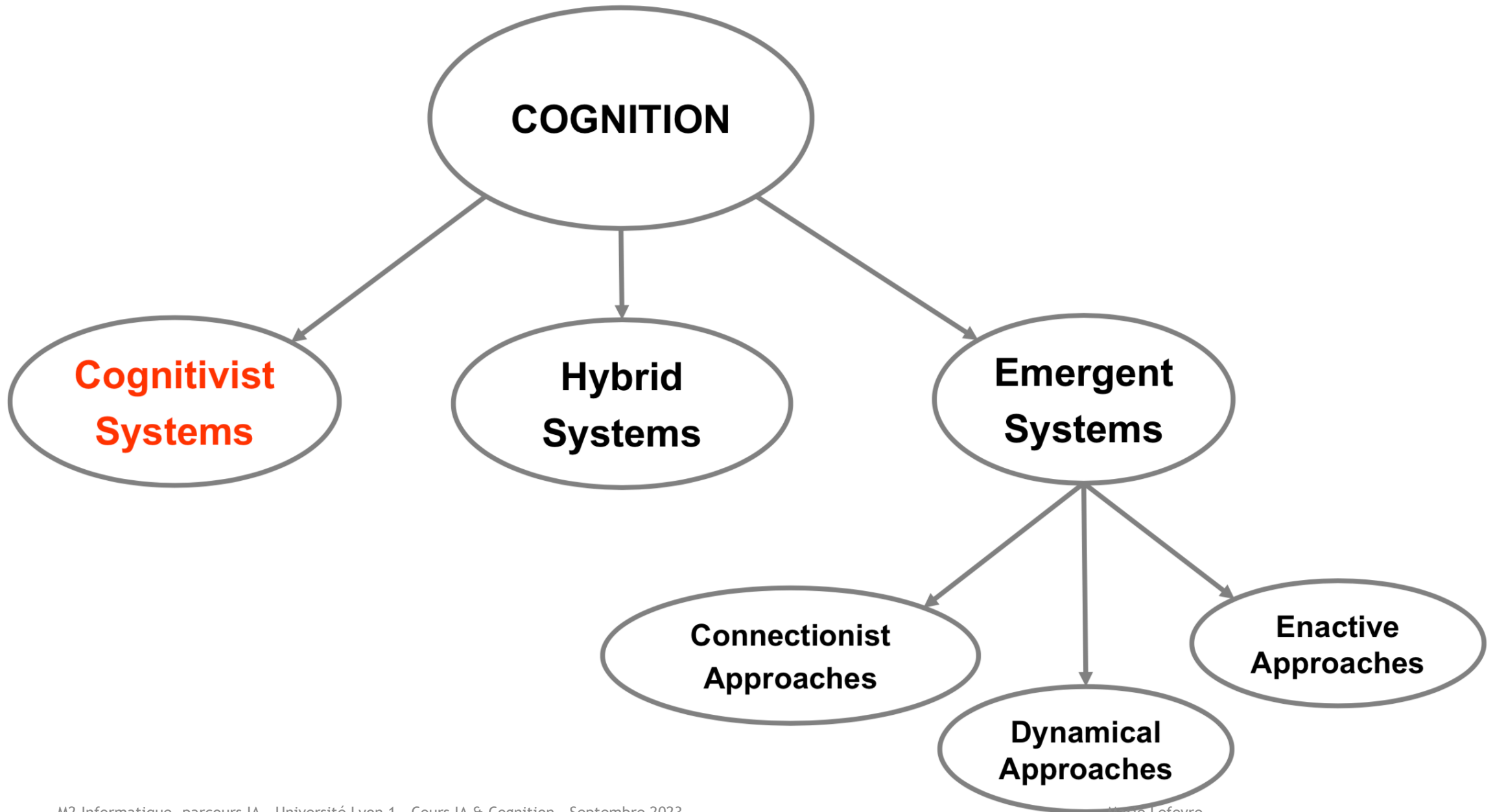
Ex : A Cognitive Agent

- Recent definition proposed by Louis Kauffman, President of the American Society for Cybernetics

"Cybernetics is the study of systems and processes that interact with themselves and produce themselves from themselves."

# Paradigmes / courants en Sciences Cognitives



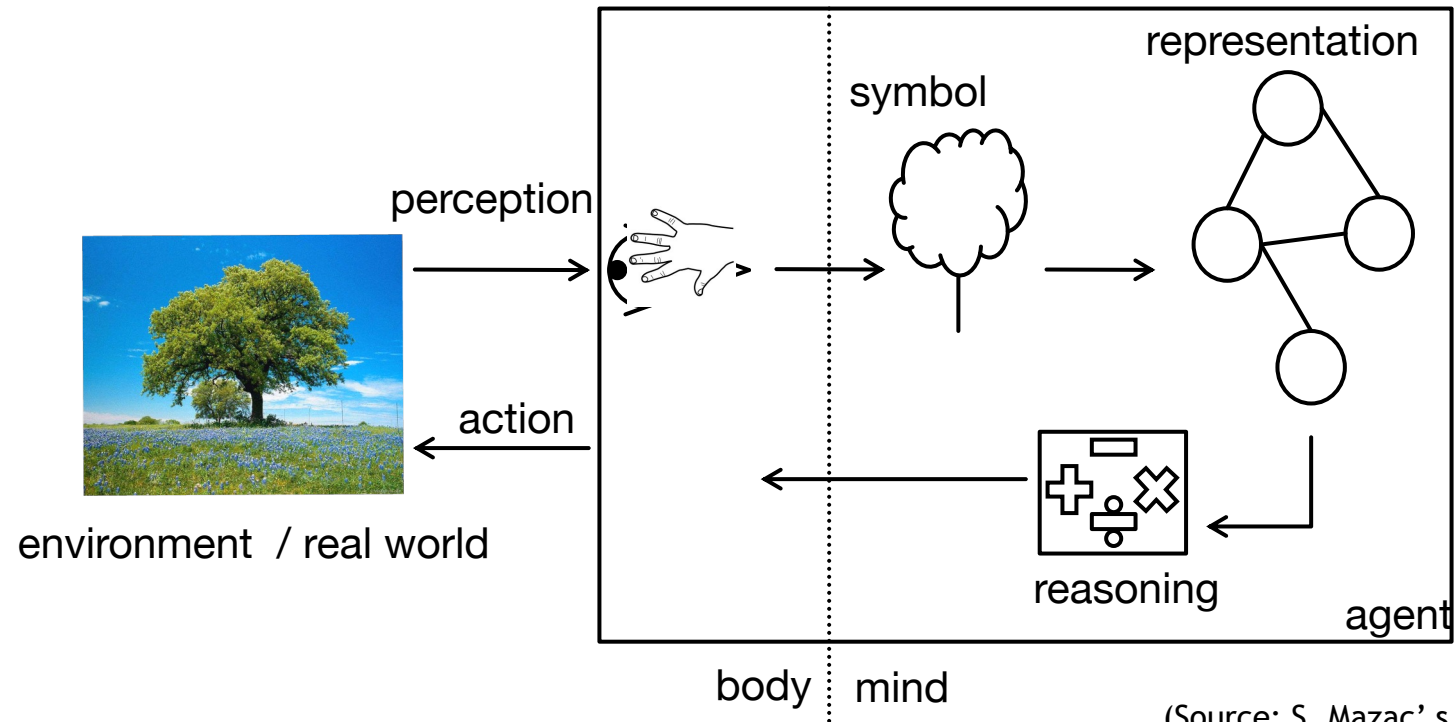




# Epistémologie positiviste : Cognitivism

- ▶ Cognition : traitement de l'information
- ▶ Représentation abstraite
- ▶ Représentation dirigée vers des objectifs précis
- ▶ Pour de nombreux problèmes, un modèle raffiné du monde peut être fourni

# Systeme cognitiviste



(Source: S. Mazac' s PhD Thesis 2015)

# Systemes cognitivistes : représentation des connaissances

- ▶ Explicite et symbolique
- ▶ Décolorée des objets externes
- ▶ Isomorphique (toute la même forme)
- ▶ Projection
- ▶ Implique une ontologie complète et accessible
- ▶ Consistant avec l'expression humaine

# General Problem Solver (Simon, Shaw et Newell 1959)

- ▶ N'importe quel problème formalisé peut en principe être résolu par GPS
  - ▶ Preuves de théorèmes, problèmes géométriques, parties d'échecs.
- ▶ Premier programme à séparer sa base de données (faits) de sa stratégie de résolution
- ▶ Nécessite une spécification des objets et les opérations applicables sur ces objets
- ▶ GPS exploite les heuristiques par une « confrontation moyens/fins » (means-ends analysis)
- ▶ GPS a résolu des problèmes simples et facilement formalisés comme les tours de Hanoï
- ▶ Pour les problèmes plus réalistes il est confronté à l'explosion combinatoire
- ▶ Le paradigme GPS a évolué vers l'architecture SOAR



# GPS : a means-ends search process

begin:

- save initial conditions as "current state"
- try to achieve all goals

how to achieve all goals:

- for each goal:
  - try to achieve the goal
- if all goals were achieved, return success

how to achieve a goal:

- if goal is already met (in current state), return success (it's achieved)
- else, for each operator:
  - if operator's "add list" contains the goal,
    - try to apply the operator
    - if successful, return success
- if no operators were successful,
  - return failure

how to apply an operator:

- try to achieve all of the operator's preconditions (treat them as goals)
- if successful (all preconditions met),
  - mark the operator as "applied"
  - add conditions in the operator's "add list" to the current state
  - delete conditions in the operator's "delete list" from the current state
  - return success
- else,
  - return failure

# Un exemple de problème



# Modélisation du problème

- ▶ Décrire ce qu'est un état du problème
- ▶ Décrire l'état initial
- ▶ Définir les **opérateurs** permettant de passer d'un état à un autre
- ▶ Disposer d'un test permettant de savoir si on a trouvé un **état but final**
- ▶ Construire l'espace des états :  
l'ensemble des états atteignables depuis l'état initial
- ▶ Construire un chemin de l'état initial à l'état final :  
une séquence d'états dans l'espace des états
- ▶ Disposer d'une fonction de coût sur le chemin :  
cette fonction associe un coût au chemin (coût calculé comme la somme des coûts individuels des actions le long du chemin)

```

problem = {
  "start": ["at door", "on floor", "has ball", "hungry", "chair at door"],
  "finish": ["not hungry"],
  "ops": [
    {
      "action": "climb on chair",
      "preconds": ["chair at middle room", "at middle room", "on floor"],
      "add": ["at bananas", "on chair"],
      "delete": ["at middle room", "on floor"]
    },
    {
      "action": "push chair from door to middle room",
      "preconds": ["chair at door", "at door"],
      "add": ["chair at middle room", "at middle room"],
      "delete": ["chair at door", "at door"]
    },
    {
      "action": "walk from door to middle room",
      "preconds": ["at door", "on floor"],
      "add": ["at middle room"],
      "delete": ["at door"]
    },
    {
      "action": "grasp bananas",
      "preconds": ["at bananas", "empty handed"],
      "add": ["has bananas"],
      "delete": ["empty handed"]
    },
    {
      "action": "drop ball",
      "preconds": ["has ball"],
      "add": ["empty handed"],
      "delete": ["has ball"]
    },
    {
      "action": "eat bananas",
      "preconds": ["has bananas"],
      "add": ["empty handed", "not hungry"],
      "delete": ["has bananas", "hungry"]
    }
  ]
}

```

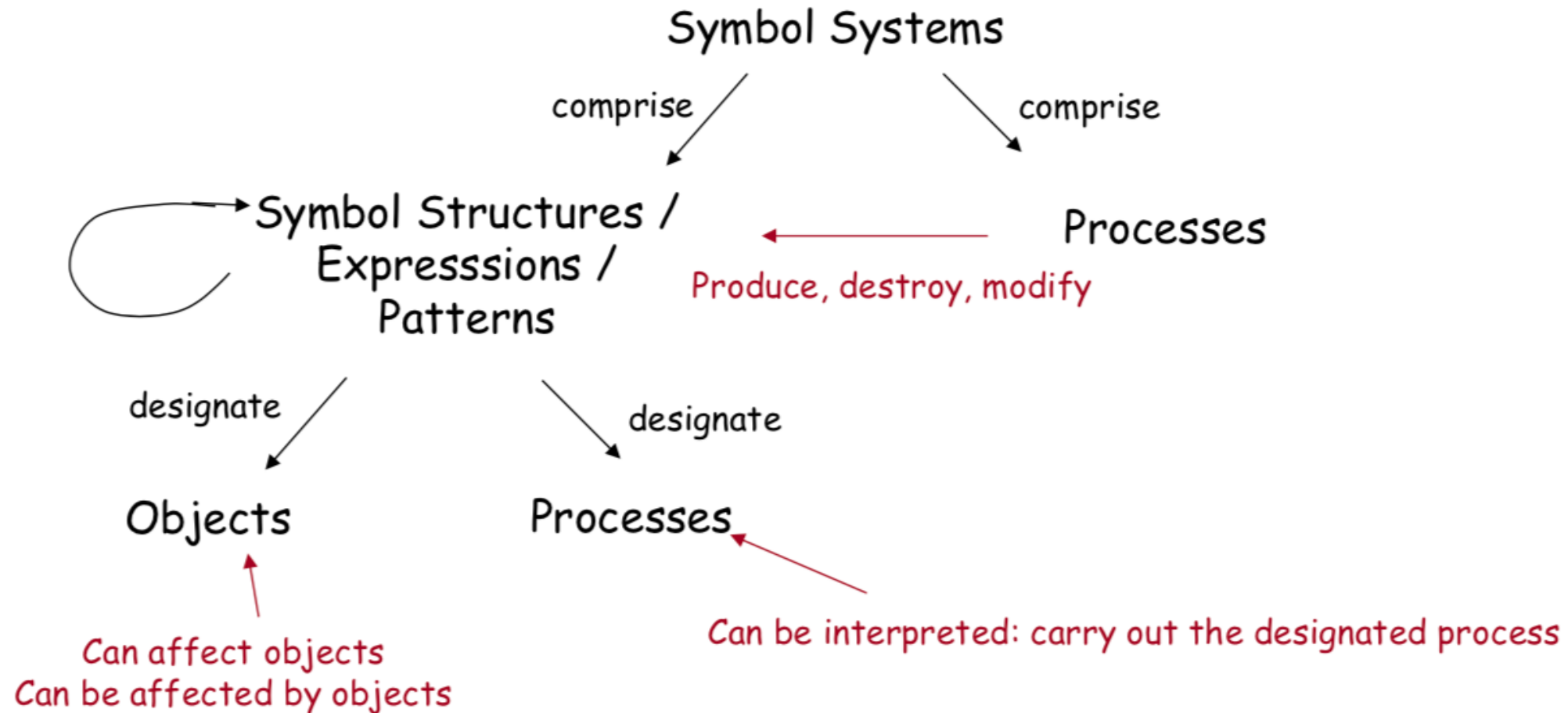




# Physical Symbol System (Newell & Simon 1975)

- ▶ Les symboles sont des entités abstraites qui peuvent être instanciées
- ▶ Un système symbolique physique a :
  - ▶ Une mémoire : pour contenir les informations symboliques
  - ▶ Des symboles : pour fournir un modèle pour faire correspondre ou indexer d'autres symboles
  - ▶ Des interprétations : pour permettre aux symboles de spécifier les opérations
  - ▶ Des capacités de composition, d'interprétabilité, de mémoire suffisante
- ▶ Le système de symboles peut être instancié mais ... le comportement est indépendant de la forme particulière de l'instanciation

# Physical Symbol System (Newell & Simon 1975)



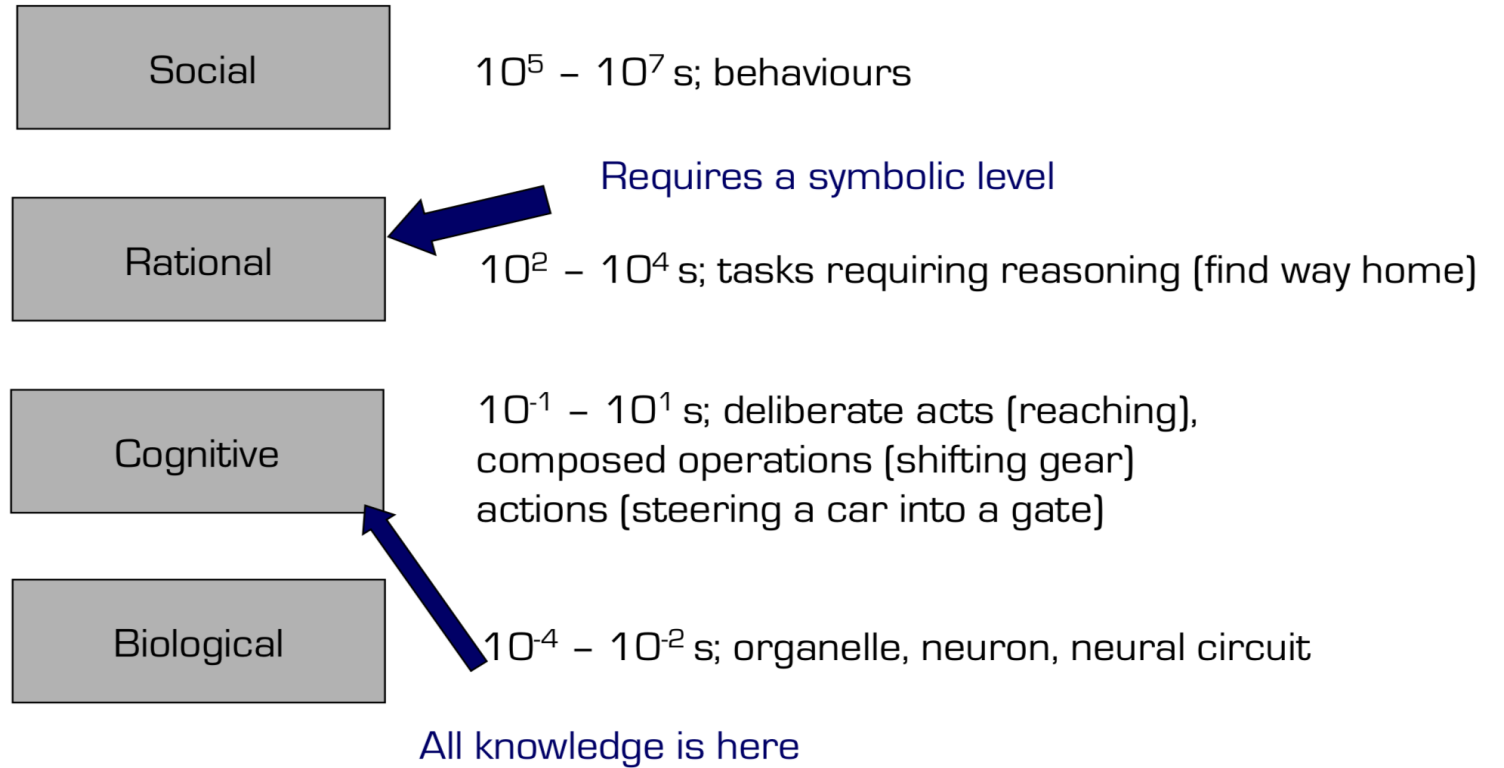
# La recherche de solution

- ▶ L'hypothèse de recherche heuristique :
  - ▶ générer et modifier progressivement des structures de symboles jusqu'à ce qu'il produise une structure de solution
  - ▶ mener une recherche efficace et efficiente
- ▶ La difficulté est d'écartier la menace omniprésente de l'explosion exponentielle de la recherche
  - ▶ Heuristique ...

# Les hypothèses

## Newell's four levels

Source: <http://www.vernon.eu>



- ▶ Un système symbolique physique a les moyens nécessaire et suffisant pour n'importe quelle intelligence
  - ▶ Qu'est-ce que cela implique pour les humains ?
  - ▶ Que l'intelligence humaine et artificielle sont équivalentes ? Et pourquoi ?

# Principes sous-jacents

- ▶ Principe de rationalité (Newell 82) : si un agent sait qu'une de ses actions conduira à l'un de ces buts, alors l'agent sélectionnera cette action
- ▶ Analyse rationnelle (Anderson 89) : le système cognitif optimise l'adaptation du comportement de l'organisme
- ▶ Théorie unifiée de la cognition (Newell 90) : nécessité d'un ensemble d'hypothèses générales pour les modèles cognitifs qui tiennent compte de toute la cognition
  - ▶ Architecture cognitive SOAR
- ▶ Un système de connaissances : peut apporter toutes ses connaissances à chaque problème
  - ▶ connaissance parfaite > utilisation complète des connaissances
  - ▶ les humains n'en sont pas encore là !

# Théorie unifiée de la cognition (Newell)

## Doit expliquer comment un esprit fait pour :

- ▶ Se comporter de manière flexible en fonction de l'environnement
- ▶ Faire preuve d'un comportement adaptatif (rationnel, axé sur les objectifs)
- ▶ Fonctionne en temps réel
- ▶ Opérer dans un environnement riche, complexe et détaillé (percevoir une immense quantité de détails changeants ; utiliser de vastes quantités de connaissances ; et contrôler un système moteur à plusieurs degrés de liberté)
- ▶ Utiliser des symboles et des abstractions
- ▶ Utiliser un langage, à la fois naturel et artificiel
- ▶ Apprendre de l'environnement et de l'expérience
- ▶ Acquérir des capacités grâce au développement
- ▶ Agir de manière autonome, mais au sein d'une communauté sociale
- ▶ Être conscient de soi et avoir un sens de soi
- ▶ Être réalisable en tant que système neuronal
- ▶ Être interprétable par un processus de croissance embryologique
- ▶ Surgir à travers l'évolution



# SOAR : architecture cognitive

- ▶ Théorie unifiée de la cognition
  - ▶ A pour but d'expliquer tous les mécanismes de tous les problèmes dans tous les domaines
  - ▶ SOAR semble montrer que c'est plausible
    - ▶ Application des principes à la cognition humaine et artificielle



ATELIER DEDIÉ

# Systemes cognitivistes : succès

## Actions in a Task

- ehow.com, wikihow.com
- Step-by-Step Instructions for Everyday Tasks



## Ontological Relations

- opencyc.org
- Very Large Encyclopedic Knowledge Base



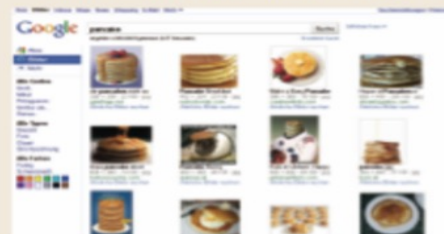
## Commonsense Knowledge

- openmind.hri-us.com
- Commonsense Knowledge from Internet Users



## Object Appearance

- germandeli.com, images.google.com
- Pictures of Products and Other Object Classes



## Object Shape

- sketchup.google.com/3dwarehouse/
- 3-D CAD Models of Household Items



## Object Properties

- germandeli.com
- Object Properties Extracted from Shopping Web sites



# Systemes cognitivistes : problèmes

- ▶ La représentation est dépendante du programmeur
  - ▶ Biais le système
- ▶ Est-il possible d'anticiper chaque éventualité lors de la conception ?
- ▶ Le fossé sémantique
  - ▶ *Symbol Grounding problem*
  - ▶ le problème de l'enracinement des symboles



ATELIER DEDIÉ

# Systemes cognitivistes : des pistes

- ▶ Comblen le fossé avec :
  - ▶ Apprentissage automatique
  - ▶ Modélisation probabiliste
  - ▶ Meilleurs modèles
  - ▶ Meilleures logiques
  - ▶ ... bref, en améliorant tout ...
- ▶ Mais reste *The Frame Problem*





ATELIER DEDIÉ


Entry Contents

Bibliography

Academic Tools

Friends PDF Preview 

Author and Citation Info 

Back to Top 

## The Frame Problem

*First published Mon Feb 23, 2004; substantive revision Mon Feb 8, 2016*

To most AI researchers, the frame problem is the challenge of representing the effects of action in logic without having to represent explicitly a large number of intuitively obvious non-effects. But to many philosophers, the AI researchers' frame problem is suggestive of wider epistemological issues. Is it possible, in principle, to limit the scope of the reasoning required to derive the consequences of an action? And, more generally, how do we account for our apparent ability to make decisions on the basis only of what is relevant to an ongoing situation without having explicitly to consider all that is not relevant?



# The Frame Problem (McCarty & Hayes 1969)

- ▶ En utilisant la logique mathématique, comment est-il possible d'écrire des formules qui décrivent les effets des actions sans avoir à écrire un grand nombre de formules d'accompagnement qui décrivent les non-effets banals et évidents de ces actions ?
- ▶ Exemple : prenons deux formules, l'une décrivant les effets de la peinture d'un objet et l'autre décrivant les effets du déplacement d'un objet
  - ▶ La Couleur (  $x, c$  ) est définie après la Peinture (  $x, c$  )
  - ▶ La Position (  $x, p$  ) est définie après le Déplacement (  $x, p$  )
- ▶ Supposons une situation initiale : Couleur ( A , Rouge ) et Position ( A , Maison )
- ▶ Avec la logique déductive, que se passe-t-il alors après l'action Peindre ( A , Bleu ) suivie de l'action Déplacement ( A , Jardin ) ?



# Axiome de cadre

- ▶ Intuitivement : Couleur ( A , Bleu ) et Position ( A , Jardin)
- ▶ Mais ce n'est pas le cas !!!
  - ▶ Dans la logique classique des prédicats, en utilisant un formalisme approprié pour représenter le temps et l'action tel que le calcul de situation (McCarthy & Hayes 1969), les deux formules autorisent uniquement la conclusion : Position ( A , Jardin )
- ▶ C'est parce qu'ils n'excluent pas la possibilité que la couleur de A soit modifiée par l' action de déplacement
- ▶ Il faudrait décrire explicitement les non-effets de chaque action : « axiomes de cadre »
  - ▶ Couleur ( x , c ) est définie après Déplacement ( x , p ) si Couleur ( x , c ) est définie avant
  - ▶ Position ( x , p ) est définie après Peindre ( x , c ) si Position ( x , p ) est définie avant

# Raisonnements non-monotones

- ▶ Axiome de cadre :
  - ▶ La plupart des actions n'affectent pas la plupart des propriétés d'une situation
  - ▶ Dans un domaine comprenant  $M$  actions et  $N$  propriétés, nous devons, en général, écrire presque  $M*N$  cadres d'axiomes
  - ▶ Solution non satisfaisante : travail fastidieux et mémoire volumineuse
- ▶ Loi d'inertie de bon sens
  - ▶ Règle générale selon laquelle une action peut être supposée ne pas modifier une propriété donnée d'une situation à moins qu'il n'y ait une preuve du contraire
- ▶ Problème de la monotonie du raisonnement logique : on ne peut que ajouter des faits
  - ▶ Quid de déplacer un objet dans un pot de peinture ?  
Il faut retirer l'ancienne couleur et ajouter la nouvelle couleur....
- ▶ Introduction de nombreux raisonnements non-monotones
  - ▶ Difficulté de fonctionner en présence d'actions concurrentes, d'actions avec des effets non déterministes, de changement continu et d'actions avec des ramifications indirectes

# Cognitivism & IA

- ▶ Logique comme support
  - ▶ Fournir un ensemble de règles qui décrivent le monde
  - ▶ **Méta-règles** : règles pour créer de nouvelles règles
  - ▶ Apprentissage de nouvelles connaissances à partir des connaissances existantes
- ▶ Problèmes
  - ▶ Apprendre au-delà des règles initiales mises dans le système
  - ▶ Apprendre des connaissances fausses
    - ▶ Révision des croyances



ATELIER DEDIÉ

# Cognitivism & IA

- ▶ Architecture cognitive : définie la manière dont un agent cognitif gère les ressources de base à sa disposition
- ▶ Dictée par ses **représentations** et leur implémentation
- ▶ Dictée par les propriétés du système cognitif :
  - ▶ Organisation et **stratégies de contrôles** (coordination / coopération, modulaire / hiérarchique)
  - ▶ Mémoire, **connaissance**, représentation (modèles du monde, représentation déclaratives / procédurales, mémoire associative, connaissance épisodique, méta-connaissance)
  - ▶ **Types d'apprentissage** (délibéré vs réflexif, monotone vs non-monotone)
  - ▶ Types de planification
  - ▶ Comportement (cohérence, saillance, adéquation, cohérence, pertinence, suffisance...)

# 4 catégories de systèmes cognitifs

## Thinking Humanly

“The exciting new effort to make computers think . . . *machines with minds*, in the full and literal sense.” (Haugeland, 1985)

“[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . .” (Bellman, 1978)

## Thinking Rationally

“The study of mental faculties through the use of computational models.”  
(Charniak and McDermott, 1985)

“The study of the computations that make it possible to perceive, reason, and act.”  
(Winston, 1992)

## Acting Humanly

“The art of creating machines that perform functions that require intelligence when performed by people.” (Kurzweil, 1990)

“The study of how to make computers do things at which, at the moment, people are better.” (Rich and Knight, 1991)

## Acting Rationally

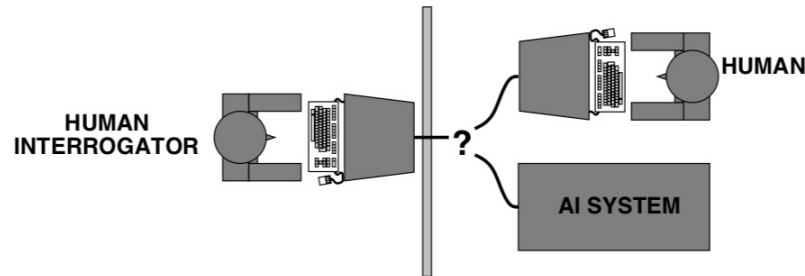
“Computational Intelligence is the study of the design of intelligent agents.” (Poole *et al.*, 1998)

“AI . . . is concerned with intelligent behavior in artifacts.” (Nilsson, 1998)

Source: Artificial Intelligence : A modern Approach (by: Russel & Norvig)

# Agir comme un humain : test de Turing

- ▶ Turing (1950) : « Computing machinery & Intelligence »
  - ▶ « Can machines think? » => « Can machines behave intelligently? »



- ▶ Prédit en 2000, une machine aura 30% de chance de duper une personne en 5 min
  - ▶ A anticipé tous les arguments contre l'IA qui ont été soulevés depuis 50 ans
  - ▶ A suggéré les composants majeurs d'une IA : connaissance, raisonnement, langage, compréhension, apprentissage
- ▶ Problème : Imitation Game est non reproductible



# Penser comme un humain :

## Sciences Cognitives

- ▶ 1960 : « la révolution cognitive » : la psychologie du traitement de l'information a remplacé l'orthodoxie dominante du comportementalisme
- ▶ Nécessite des théories scientifiques sur les activités internes du cerveau
  - ▶ Quel niveau d'abstraction ? Est-ce des connaissances ou des connexions ?
  - ▶ Comment valider ?
    - ▶ Prédire & tester le comportement de sujet humain (top-down)
    - ▶ Identifier à partir de données neurologiques (bottom-up)
- ▶ Différentes approches en IA :
  - ▶ À peu près
  - ▶ Sciences cognitives
  - ▶ Neurosciences cognitives

# Penser rationnellement : Laws of thought

- ▶ Normatif ou prescriptif plus que descriptif
- ▶ Aristote :
  - ▶ Quels sont les arguments/processus de pensée corrects ?
  - ▶ Plusieurs écoles grecques ont développées diverses formes de logique :
    - ▶ notation et règles de dérivation pour les pensées
- ▶ Exploitation des mathématiques et la philosophie pour l'IA moderne
- ▶ Problèmes :
  - ▶ Tous les comportements intelligents ne sont pas médiatisés par le raisonnement logique
  - ▶ Quel est le but de penser ? Quelles pensées dois-je avoir ?

# Agir rationnellement

- ▶ Comportement rationnel : faire les choses « justes », « bonnes »
- ▶ Qu'est-ce qu'une chose « juste »
  - ▶ Ce qui amène à maximiser l'accomplissement d'un but,
  - ▶ à obtenir l'information souhaitée
- ▶ Il n'est pas nécessaire de penser (i.e. cligner des yeux) mais penser peut aider à avoir une action rationnelle

# Pour aller plus loin

- ▶ De la lecture :
  - ▶ Stuart Russell & Peter Norvig , Artificial Intelligence : A Modern Approach,
  - ▶ David Vernon, Artificial Cognitive Systems, MIT Press, 2014.
  
- ▶ Sources pour construire ce cours
  - ▶ Les cours de Salima Hassas et Amélie Cordier

# Les ateliers

- ▶ Chaque atelier a sa propre thématique, relative à un article de référence
  - ▶ 4 séances, 2 groupes par séance, **1 groupe = 2-3 personnes**
- ▶ Travail à faire par le groupe
  - ▶ **15/09 : choix de 3 thématiques triés par ordre de préférence, avec noms du groupe et contraintes de dates**
  - ▶ **18/09 : affectation des thématiques et de la date**
  - ▶ Approfondir le sujet en cherchant d'autres ressources éventuellement
  - ▶ Le jour J
    - ▶ 20 min de présentation oral sur le sujet
    - ▶ 20 min de débat sur le sujet avec la salle
    - ▶ 5 min de synthèse
  - ▶ A J+7 : fiche de synthèse de 1 à 2 pages max (sur l'atelier : sur le contenu, les débats, etc. )
- ▶ Le groupe est noté sur la présentation, l'animation de l'atelier (qualité de la discussion) et la fiche de synthèse

# Les thématiques

- ▶ Chinese Room
- ▶ The Frame Problem
- ▶ The Symbol Grounding Problem
- ▶ Commonsense problems
- ▶ MetaCognition
- ▶ The SOAR (State Operator and Results) Architecture
- ▶ Embodied cognition
- ▶ Intrinsic Motivation
- ▶ Social Robotics
- ▶ **Autres ?**