Histoire, épistémologie et concepts fondamentaux des sciences cognitives

Jérôme Sackur
DEC / LSCP

Des sciences naturelles de l'esprit

- Sciences de la nature / sciences de l'esprit (Dilthey)
  - Objets différents
  - Méthodes différentes: expliquer / comprendre
  - Lois causales / restitution d'une totalité

- Les sciences cognitives, c'est le contraire:
  - Appliquer les méthodes des sciences de la nature à l'esprit

Charybde et Scylla

Comment traiter de l'esprit *de manière naturaliste* sans faire disparaître l'esprit?

- Charybde: le béhaviorisme
- Scylla: le « spiritualisme »

L'esprit résisterait-il toujours?

Pourquoi?

- Non reproductibilité (histoire collective, psychologie individuelle) (Kant)
- Fluence, évanescence (Kant)
- Qualitatif et non quantitatif (Bergson, Nagel: les *qualia*)
- Complexité

Le Béhaviorisme

- J. Watson (1913): « Psychology as the behaviorist views it »
- B. F. Skinner (1938) *The Behavior of organisms*
- La psychologie est la science du *comportement.*
- Prévoir et contrôler le comportement
- Comment se passer du contenu de la boîte noire?

« Albertine va au cinéma parce qu'elle veut se changer les idées »

Toutes nos explications naïves font appel à des entités mentales: des *représentations*

- Théorie de l'apprentissage total:
- le comportement de *X* au temps *t* est prédictible par:
  - L'ensemble des stimuli
  - L'histoire totale de son conditionnement avant *t*.

Deux formes de conditionnements

- Le conditionnement pavlovien
  - avant conditionnement A^0, US^+
  - conditionnement: A+US
  - après conditionnement A^+, US^+
- Le conditionnement opérant:
  - Avant conditionnement: répertoire d'opérants: (<A,p_{t_o}>, <B,q_{t_o}>, <C,r_{t_o}>, <D,s_{t_o}>,...)
  - Conditionnement: A -> +
  - Après conditionnement (<A,p_{t_1}>, <B,q_{t_1}>, <C,r_{t_1}>, <D,s_{t_1}>,...) avec $\mathbf{p_{t\_1} > p_{t\_0}}$
- L'histoire passée d'un organisme explique son comportement présent

La superstition chez le pigeon

A pigeon is brought to a stable state of hunger by reducing it to 75 percent of its weight when well fed. It is put into an experimental cage for a few minutes each day. A food hopper attached to the cage may be swung into place so that the pigeon can eat from it. A solenoid and a timing relay hold the hopper in place for five sec. at each reinforcement.

If a clock is now arranged to present the food hopper at regular intervals *with no reference whatsoever to the bird's behavior*, operant conditioning usually takes place. In six out of eight cases the resulting responses were so clearly defined that two observers could agree perfectly in counting instances. One bird was conditioned to turn counter-clockwise about the cage, making two or three turns between reinforcements. Another repeatedly thrust its head into one of the upper corners of the cage. A third developed a 'tossing' response, as if placing its head beneath an invisible bar and lifting it repeatedly.

The experiment might be said to demonstrate a sort of superstition. The bird behaves as if there was a causal relation between its behavior and the presentation of food, although such a relation is lacking.

Skinner, 1948

Mérites et limites du béhaviorisme

- Les chiens (et les humains) ne *prévoient* pas l'arrivée de la nourriture, ils sont conditionnés
- Les pigeons (et les humains) peuvent devenir superstitieux
- Le langage est le résultat d'un entre-conditionnement complexe. BF Skinner *Verbal Behavior* (1958).

Le Béhaviorisme est:

- Anti-représentations
- Naturaliste, objectiviste
- Non-réductionniste
- Empiriste

La chute du béhaviorisme

- Démenti par les faits: observation learning, latent learning, etc.
- Chomsky (1959): « A review of Skinner's *Verbal Behavior* »: explication non plausible compte tenu de la *pauvreté du stimulus*

Opposition de deux paradigmes:

- Le béhaviorisme est peu plausible mais non « falsifiable »; multiplication des hypothèses *ad hoc*. (sub-vocalisation; connaissance *exhaustive* du passé de l'organisme).
- Le cognitivisme s'appuie sur des intuitions fortes

Le *coup de grâce*

- La théorie moderne du conditionnement fait appel aux représentations mentales!
- Théorie du conditionnement fausse: la contiguité spatio-temporelle ne fait pas le conditionnement. Exemple (le blocking):
- A et B deux stimulus neutres; US un stimulus inconditionné

    1. A^0, B^0, US^+
    2. A+US
    3. A+B+US
    4. B^0 ou B^+ ?

- Ce qui conditionne: la différence entre le monde et sa représentation

Les représentations

Le paradigme cognitif est **mentaliste**:

- il y a des niveaux de descriptions des organismes ou systèmes qui impliquent des états ou représentations mentales
- Ces représentations ne sont pas éphiphénoménales: elles ont une efficacité causale
- Les représentations ont une contrepartie matérielle...
- ... mais ne se réduisent pas à du matériel

➔ “Monisme ontologique, dualisme nomologique”

Le point d'appui formel: logique et informatique

- Actuellement, sciences cognitives ~ cerveau
- Mais rien n'aurait été possible sans les sciences formelles (logique, informatique théorique).

Les grandes étapes:

- Du raisonnement au calcul (Frege, 1879: *Idéographie*)
- Du calcul au mécanisme:
  - La machine logique (Turing, 1936)
  - L'ordinateur (Von Neumann, 1945)
- Du mécanisme au cerveau (McCulloch et Pitts, 1943)

Le raisonnement comme calcul: langage et système formels

- Frege, 1879: un langage calculatoire pour des preuves « sans lacune ».
- Exemple, le *Modus Ponens*: « S'il fait beau, je sors le parasol; or il fait beau; donc je sors le parasol ».

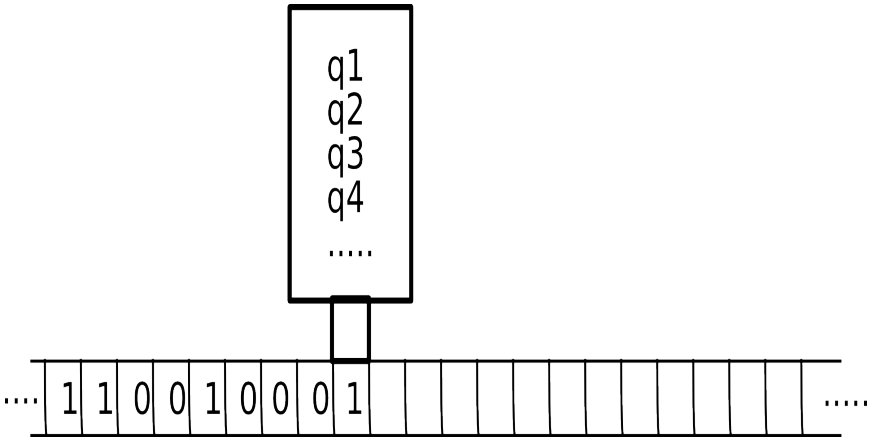
$$\frac{\chi \qquad \chi \rightarrow \gamma}{\gamma}$$

- L'application des règles logiques est vérifiable sans recours à l'intuition.
- Penser est une forme de calcul mécanisable.

La machine logique: Turing, 1936

- « Nous pouvons comparer un homme en train de calculer un nombre réel à une machine capable de se trouver seulement dans un nombre fini d'états q_1, q_2, … q_R que l'on appellera « configurations-m ». La machine dispose d'un « ruban » (analogue à une feuille de papier) qui défile en elle et qui est divisé en sections (appelées « cases ») dont chacune est susceptible de recevoir un « symbole ». A un moment quelconque, il n'y a qu'une case, disons la r-ième, portant le symbole S(r), qui soit « dans la machine ». Nous pouvons l'appeler la « case inspectée ». Le symbole inscrit dans la case inspectée est le « symbole inspecté ». Le « symbole inspecté » est le seul dont la machine soit, pour ainsi dire, « directement consciente ». (…) Le comportement possible de la machine à un moment quelconque est déterminé par la configuration-m q_n et le symbole inspecté S(r). [Ce comportement se borne à l'écriture d'un symbole, son effacement, un changement de configuration-*m*, et/ou un déplacement de la bande.] (…) Je soutiens que ces opérations comprennent toutes celles que l'on utilise pour le calcul d'un nombre. »

- Tout ce qui est calculable est turing-calculable (thèse *spéculative* de Church-Turing).
- Il existe une machine de Turing universelle (théorème)



L'ordinateur de von Neumann

- Pas un simple calculateur...
- ... mais approximation **finie** d'une machine de Turing universelle
- Problèmes d'architecture: “un cpu, une mémoire”? Plusieurs cpu?...
- Problème physique de la « discrétisation »: qu'est-ce qui fait qu'un état ou entité physique est un *symbole*?
- Règle à calcul (analogique) contre boulier (discret).

L'ordinateur pour le paradigme cognitif: un système physique symbolique

A physical symbol system consists of a set of entities, called symbols, which are physical patterns that can occur as components of another type of entity called an expression (or symbol structure). Thus a symbol structure is composed of a number of instances (or tokens) of symbols related in some physical way (such as one token being next to another). At any instant of time the system will contain a collection of these symbol structures. Besides these structures, the system also contains a collection of processes that operate on expressions to produce other expressions: processes of creation, modification, reproduction, and destruction. A physical symbol system is a machine that produces through time an evolving collection of symbol structures. Such a system exists in a world of objects wider than just these symbolic expressions themselves.

A. Newell, H. Simon « Computer Science as empirical inquiry », 1976

L'ordinateur: une métaphore?

We wish to emphasize that we are not using the computer as a crude analogy to human behavior—we are not comparing computer structures with brains, nor electrical relays with synapses. Our position is that the appropriate way to describe a piece of problem-solving behavior is in terms of a program: a specification of what the organism will do under varying environmental circumstances in terms of certain elementary information processes it is capable of performing. This assertion has nothing to do—directly—with computers. Such programs could be written (now that we have discovered how to do it) if computers had never existed.[2]

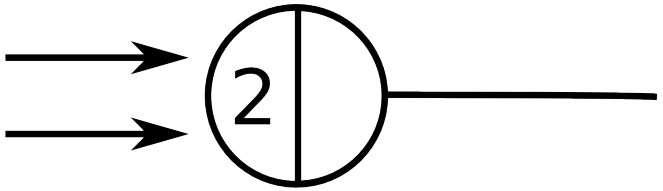
Newell, Shaw, Simon, 1958: « Elements of a theory of human problem solving »

Niveaux de description

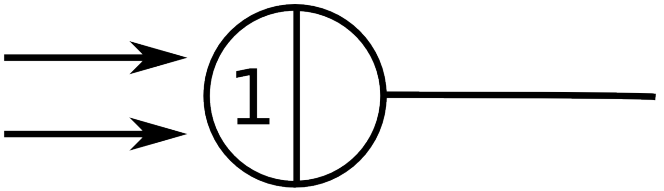
- Les ordinateurs sont importants:
  - Parce qu'on y met des programmes
  - Parce qu'ils existent
- Le fonctionnalisme: la même réalité cognitive est réalisable de diverses manières. (Réalisabilité multiple).
- Les trois niveaux de David Marr (1978, *Vision*):
  1. Computationnel (but des caculs).
  2. Algorithme et représentations (comment réaliser les calculs).
  3. Implémentation matérielle.
- Exemple du thermostat.

Le cerveau comme machine logique: McCulloch et Pitts, 1943

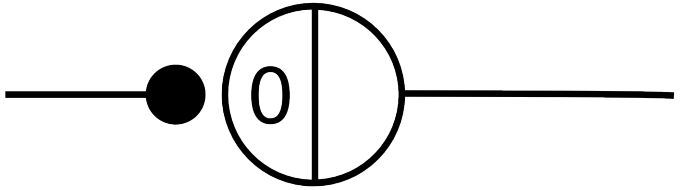
- Les neurone formels:
  - machines « tout ou rien »
  - en réseaux logiques



Connecteur "et"



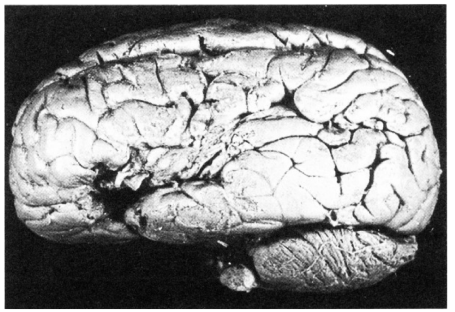
Connecteur "ou"



Connecteur "non"

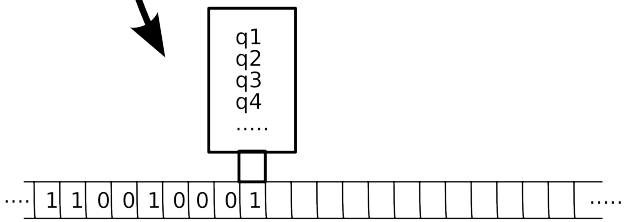
Un résumé?

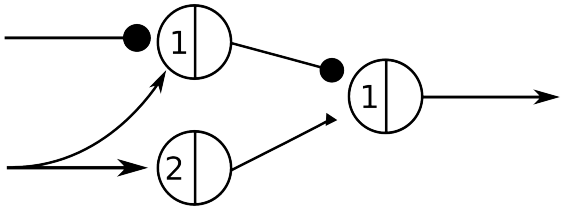
Systèmes physiques



Machines logiques

Réseaux de neurones





Objets mathématiques

Les représentations en pratique

- On peut modéliser des réalités mentales...
- ... Tester les modèles expérimentalement
- ... Simuler et comparer avec les données expérimentales.
- Le comportement est le *moyen* de l'étude du mental, pas la *fin.*

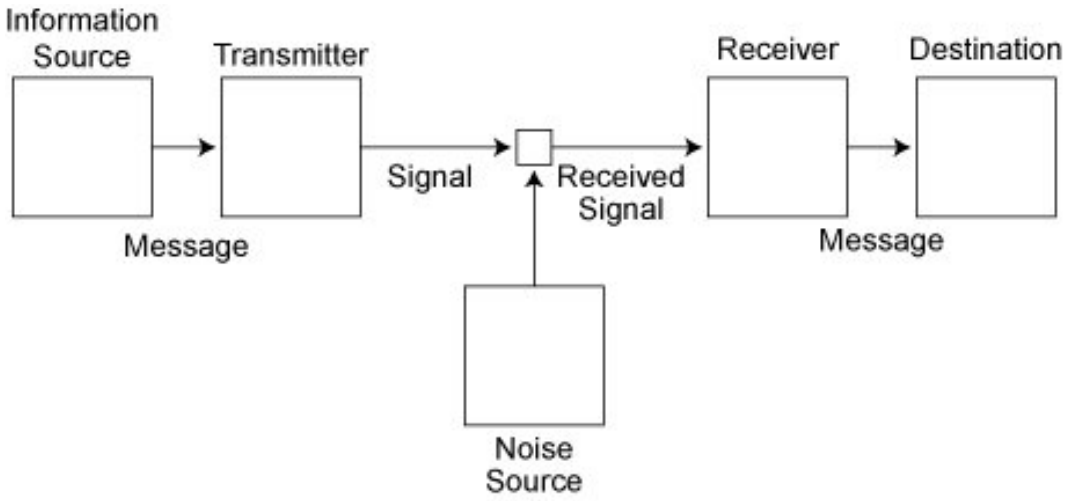
- L'esprit naturalisé ... et non dénaturé

Remarque 1: le cerveau

- Sans l'application de la théorie cellulaire au système nerveux (Golgi, Ramon y Cajal...) rien n'aurait été possible.
- Le cerveau n'est pas une masse uniforme
- On peut détailler des circuits neuronaux (Sherrington 1910)
- L'information circule de manière discrète: potentiels d'action (Du Bois Reimond 1849 – Hodgkin & Huxley 1952)
- La neuropsychologie (Broca, 1860) montre qu'il y a une inscription fine des représentation mentales.

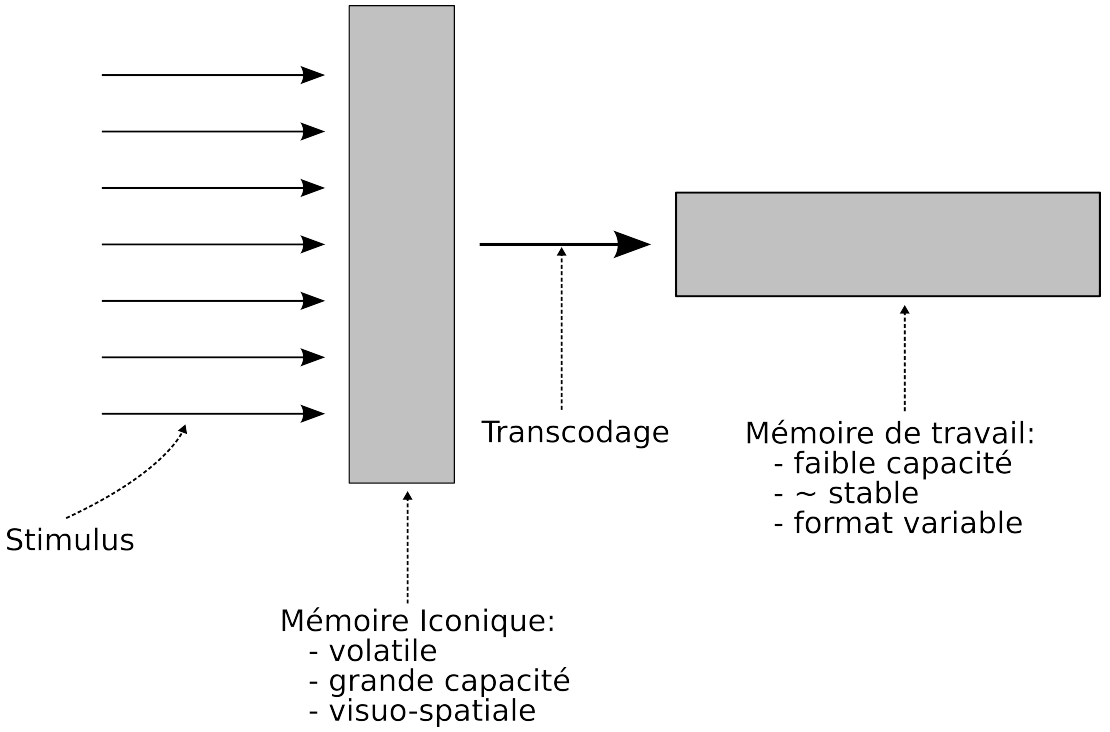
Remarque 2: le traitement de l'information

- Représentations symboliques / information (Shannon, 1948)
- Quantification de l'information sans rapport à la signification
- L'information définie comme réduction d'incertitude



Remarque 2: le traitement de l'information

- Les organismes sont des « informations processors »
- Broadbent (1958) *Perception and Communication*
- Sperling (1960) « The information available in brief visual presentations »

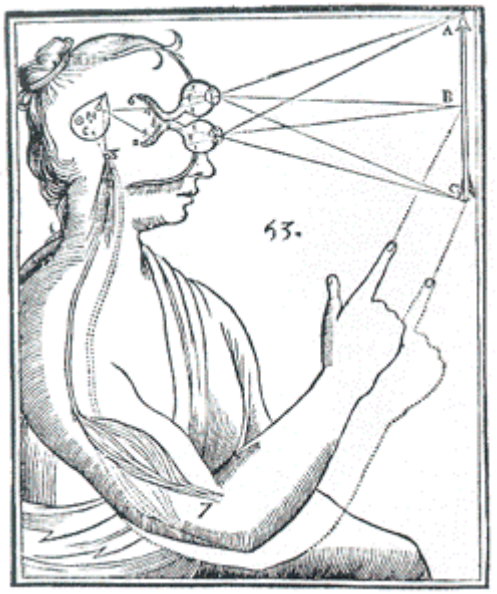


Remarque 3: A quoi sert la logique?

- Au raisonnement
- Au langage (Chomsky)
- A la perception (Helmholtz, 1865): percevoir est interpréter les signes sensoriels au moyen d'inférences *inconscientes*

Remarque 4: Trois « mécanismes »

- Le mécanisme de l'automate (Descartes, Pavlov, Watson)
- Le mécanisme flexible (Skinner)
- Le mécanisme des machines logiques: le cognitivisme



Histoire récente

- Les points d'équilibres des sciences cognitives:
- Psychologie (1960->1970)
- Intelligence artificielle (1970s).
- Limites du paradigme « classique »:
  - Ambition généraliste démesurée: Newell et Simon « general problem solver »
  - Indifférence:
    - au stimulus: de la représentation à l'information
    - au contexte biologique de l'évolution de l'esprit
  - Excès « logicistes » (computationalistes)
- Actuellement (fin 1990 ->): imagerie, neurophysiologie.

Quelques questions et directions

- L'inné et l'acquis, la modularité
- La conscience et la subjectivité
- La génétique
- L'ouverture aux sciences sociales (économie, théorie de la décision)

L'innée et l'acquis, la modularité

- Le béhaviorisme suppose que tout est acquis
- Le cognitivisme a une tendance innéiste:

➔ Thèse modulariste: l'esprit n'est pas une faculté générale, mais est décomposable

- Neuropsychologie: dissociations. Ex: aphasies de Broca / de Wernicke
- Linguistique, Chomsky:
  - pauvreté du stimulus
  - Faculté d'apprentissage spécialisée

La modularité

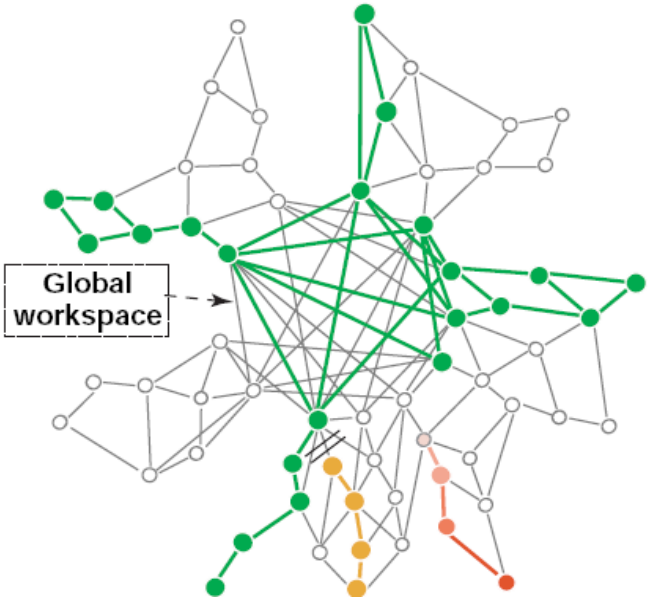
- Philosophie, Fodor (1983). Les modules sont:
  - Domaines spécifiques
  - Encapsulés
  - Automatiques et fonctionnent en parallèle
  - Cognitivement impénétrables
- Evolutionnisme (Cosmides & Tooby, 1992):
- Problème du recyclage neuronal

La conscience et la subjectivité

- Point de départ oublié, refoulé: une science de l'esprit?
- Victoire du cognitivisme: le mental n'est pas seulement le subjectif
- Mais que faire du subjectif? Le *je-ne-sais-quoi* ou le *what-it-is-like* (to be a bat, Nagel, 1974).
- Deux concepts de conscience (Block, 1995):
  - Conscience d'accès (rapportabilité, contrôle, planification, mémorisation)
  - Conscience phénoménale: qualité subjective de l'expérience

La conscience et la subjectivité

- Un modèle de l'accès: le global neuronal workspace (Baars, 1989, Dehaene, Changeux...)
- Sont accessibles des représentations partagées par une majorité de processeurs centraux



La conscience et la subjectivité

- Comment résoudre la question de la phénoménalité?
  Illusion cognitive (O'Regan, Dehaene):
  - Les représentations conscientes sont éparses
  - Le monde est une mémoire externe
- Verrou méthodologique forcé par l'imagerie? (Block)

A suivre

- CO 1 « Introduction à la philosophie de l'esprit »
- Journal Club et atelier théorique.