Computer and consciousness

what does it mean : to be conscious of something?

(ECAP -Montpellier, june 2008)

Framework

- Introduction

- A short glance at history of philosophy

- Biological and artifical representations of brain

- Toward a computer's consciousness

- Conclusion

Introduction

- We shall begin with a short recall of the history of the concept of consciousness as it takes place in the history of philosophy.

- Then we shall try to say what biological and technological approaches have brought in the debate.

- Finally, we shall show what actually may be done to simulate some kind of consciousness in recent artificial systems and why it is not, for a philosopher, sufficient to explain the ego.

1

A short glance at the history of philosophy

A cartesian discovery

- Conciousness is not a concept available in the Ancient Times (Plato or Aristotle speaks of « soul », but not of « consciousness »).

- The notion of consciousness appears only in the first half of seventeenth century when Descartes (Meditations, II) proves that the words « ego cogito, ergo sum » (I think, so I am) were always true, so he takes consciousness as one of the basis of science (the other being God).

- What is this ego, conscious of himself? He is :

- First, a « substance », which only means, at first, an exception to the doubt (the doubt being the method Descartes was using to reach some truth).

- Secondly this « ego » is a mind or a reason, whose essence has to be thought of, actually, as completely distinct from the essence of matter.

Kantian Form

- It appears very soon that Descartes'view of a substantial (but non extensive) mind was indeed very problematic :

- This non extensive mind looks like a kind of « ghost » in the body's machinery ;

- And as Kant said, nobody can take a pure nor empirical intuition of this mysterious « substance ».

- There is no pure intuition of consciousness,because for Kant, the only pure intuitions are space and time.

- There is no empirical intuition of consciousness, because the empirical ego is always changing.

- Moreover, as, for Kant, there is no intellectual intuition at all, this philosopher concluded that the « I », who accompanies all our representations, is a pure form.

Husserl's theory

- Let us explain now the approach of the german philosopher Edmund Husserl who attributed to consciousness the property of intentionality (already present in Brentano's work).

- When Descartes or Kant said that the ego has some representation of the world, they seem to suppose that this entity was purely passive.

- On the contrary, the concept of « intentionality » explains that the « I » is very active : for Husserl, every consciousness is always a consciousness *of something*, and so the ego has also feelings, projects, trusts about the world, etc..

- So the world itself is not a box in which the ego would stand as an object among others. The world is essentially, for the ego, a series of occasions of actions.

Ryle's behaviorism

- Let us now explain briefly the viewpoint of behaviorism.

- From Descartes to Husserl (and even to Bergson), a spiritualist view of mind has been prevalent in philosophy : mind (even reduced to a pure form) was supposed to be independent of body.

- By a detailed study of logical powers of mental concepts (knowing, listening, suffering, getting an image, and so on), the english philosopher Gilbert Ryle put forward that consciousness data were just a myth.

- According to him, one could explain such concepts in a satisfactory manner by refering to aptitudes for action, aquired tendencies, styles of behaviour, etc., without resorting to some improbable « inner life ».

- This behaviorist program was a kind of reductionism, which did have many revivals afterwards, until biology brings some new light on the subject.

2
Biological and artifical representations of brain

Biological viewpoint

- During the 20th century, progress in biology of mind shew there was actually a « conscious brain » (Rose,1972).

- After the works of Golgi, Cajal, Sherrington, Hebb and others, the structure of the cortex brain became to be better understood and revealed a complex discrete neural network with both electric and chimical shared and multi-distributed processes.

- So some authors like Eccles, Rosenblueth, Koestler and Smythies began to think, as one said, « beyond reductionism ».

- It means that those scientists contest the idea that it would be always possible to reduce the « inner life » of consciousness to some well identified behaviours.

- Of course it does not mean that consciousness could exist as an entity that would be distinct from its biological frame (as Descartes trusted it) but only that we cannot establish a clear bijection between the brain and the mind.

Progress in technology

- In the same way, progress in technology rose some new kinds of models of mind.

- In the context of Wiener's cybernetics, for instance, K. Deutch (1948) distinguished two kinds of messages in the brain :

- First, messages moving through the system of brain in consequence of its interaction with the outside world (which he called « primary messages »).

- Secondly, messages about changes in the state of parts of the system of brain (which he called « secondary messages »).

- So, he introduced a concept of consciousness as a collection of internal feedbacks of those special « secondary messages ».

Artificial Intelligence approach

- After 1956, the development of Artificial Intelligence (A.I) suggested that machines could perform a process which is very analogous to consciousness, because they not only had operational programs, but also knew of them : For instance, they can be aware of their subroutines and able to reorganize them if needed.

- M. Minsky wrote in 1968 that it would not be very hard, technically, to put in machines some features that we would have to call self-awareness of a kind.

- Then, advances in A.I, around the 1980's, were at the origin of a lot of researches about language of thought, modularity of mind (Fodor), computational models of reasoning, and finally,meta-level inference and consciousness

Further debates

- With Dennett (1991), Husserl's intentionality came back into fashion and everybody started speaking of «intentional artificial systems» that perhaps could explain consciousness.

- Big debates about Turing's Test (Davidson), mental models (Johnson-Laird), biology and intentionality (Searle), society and computer models of mind (Boden) appeared in those years.

- But even with the help of the strange loops of Hofstadter (1979), consciousness remained a mystery or even a myth (Engel, 1994).

- During all those years, many other reductionnist approaches try to naturalize consciousness properties as they appear in language (Chomsky) intentionality (Fodor) or thought (Dretske), without a complete success.

A physics of mind?

- At the end of the 20th century, models coming from physics of matter inspired Roger Penrose, Tom Siegfried or August Stern, who propose to explain the properties of consciousness by the help of quantum mechanics.

- This situation, which would have horrified a philosopher like Descartes, may be understood for two reasons :

- First, for quantum physics, particles of matter are associated with waves, which are not exactly matter but rather probability information.

- Secondly, some principles of quantum mechanics (for instance, Heisenberg's one) show that those particles of matter have (like higher level organisms) some kind of undeterminist behaviour.

- The fact remains that it is difficult to speak of « consciousness» when there exists no brain nor even some kind of biological frame (Edelman, 1992). So, generally, biologists are not convinced of those theories.

Toward a computer's consciousness?

- What do biologists think now?

- Today, there are still many different theories of consciousness among biologists (Missa, 1993).

- A first approach consists of thinking that consciousness comes from the structure and working of the brain (Edelman's model).

- A second view would be that there exists high level structuring forces, which are cognitive forces controling all the physical movements that may be observed (Sperry-Eccles' model).

- A third approach comes to say that consciousness is an emergent property of the brain, viewed as a self-organizing system capable of self-observation (Maturana-Varela, 1994; Atlan, 1995).

- But there exists a lot of other theories…

- As I cannot speak of all those approaches, I just want to say some words about Edelman's models and some artificial attempt to realize it or to go a little beyond.

An example : Edelman's theory

- Edelman's theory is grounded on three principles :

- 1. «Neurons that fire together wire together»

- 2. There are clusters of neurons associated to connected stimuli, which make couples of maps.

- 3. There exists re-entries between those maps.
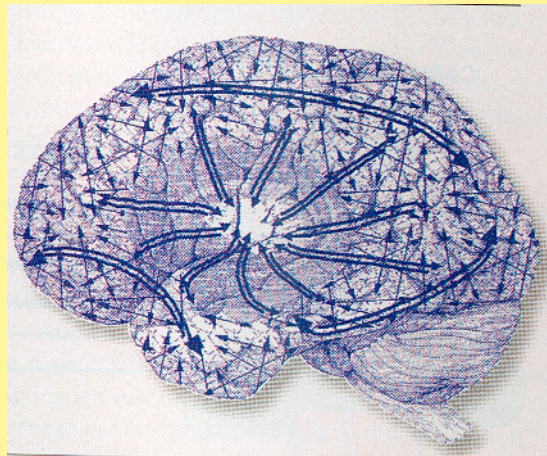
Lower level of consciousness

- This simple theory explains quite well what we can call « primary consciousness » (i.e. the fact of being conscious that things are present in the world, and the fact of having mental images of that).

- The problems are :

- What about higher states of consciousness : feeling of a person, its present, its future, and so on?

- What about consciousness of the world when symbolic systems (language, music, visual arts) come into play?

What can we make now?

- On the basis of Edelman's view, today, we can only build artefacts capable of learning from the environment.

- Their main characteristics are :

- Behavior without programming

- Possible visual categorization (for instance, catching square objects, avoiding round objects, etc.)

- Learning and conditioning without instructions.

- An Example of such a simulation artefact is the neuromecanics which name is «DARWIN ».

Darwin
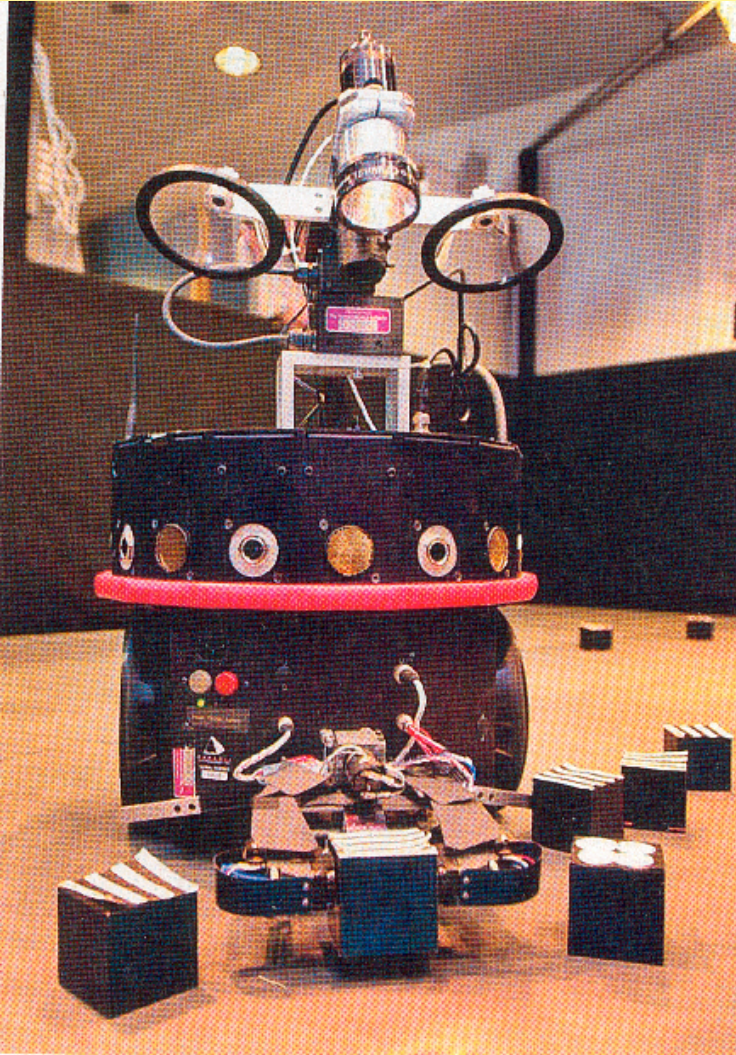
This is a view of Darwin (after the french magazine « La Recherche », whose recent issue is devoted to « Consciousness »)
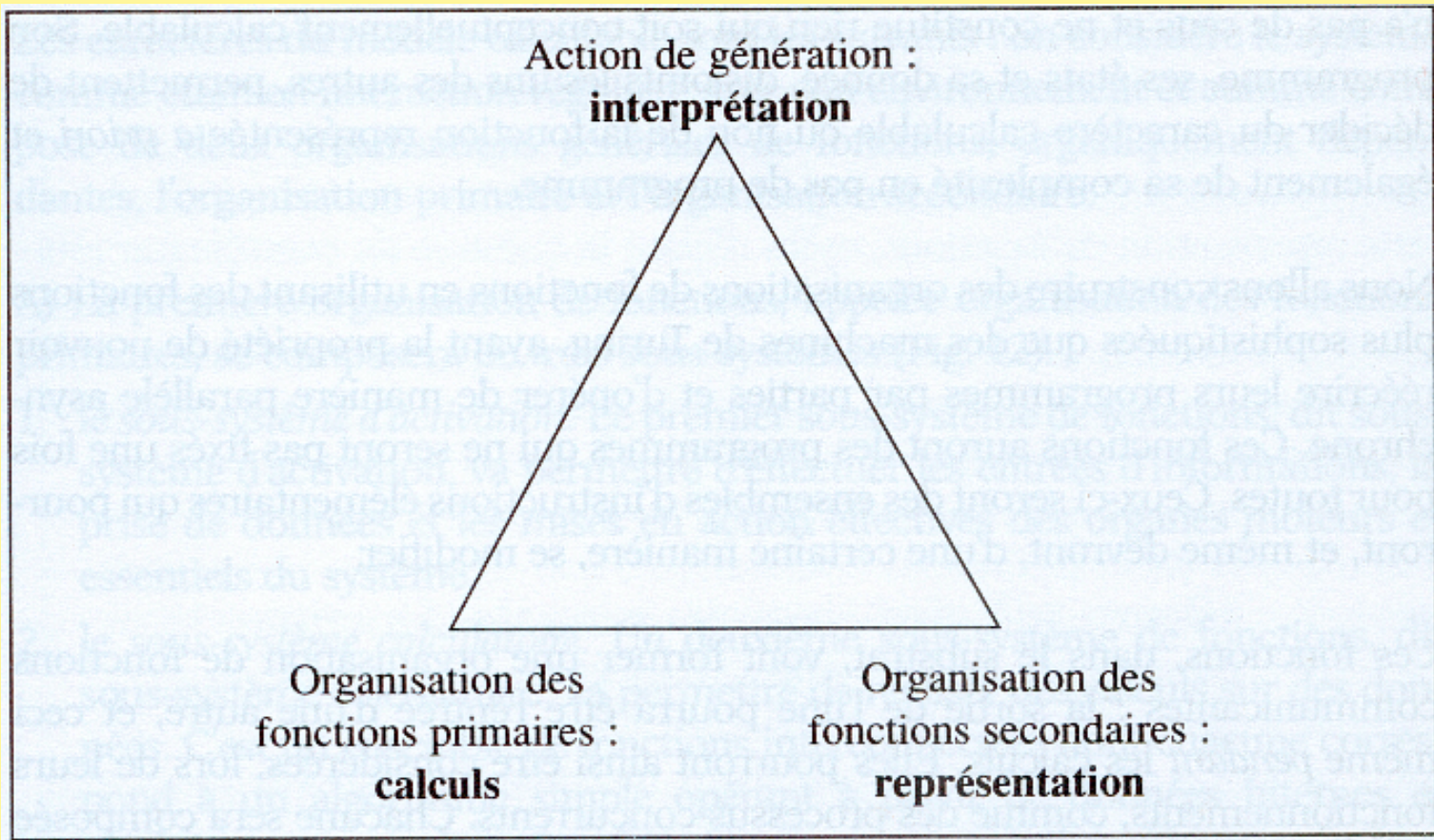
(n°30, février 2008)

Darwin can recognize simple objects in its environment.

Cardon's research

- Can we go further?

- In the beginning of years 2000, french computer scientist Alain Cardon designed a system which was, according to him, capable of generating « facts of consciousness ».

- This system would have to integrate three levels : perception, representation and interpretation of the world.

- So it is organized in three parts :

- A subsystem of primary functions (computing the elements of perception)

- A subsystem of secondary functions (representing these data)

- A subsystem of generation of conscious facts (interpreting all the previous information in the context of different environments)

Main design

Action de génération :
interprétation

Organisation des
fonctions primaires :
calculs

Organisation des
fonctions secondaires :
représentation

An artificial conscious system?

- According to Cardon, consciousness would be attached to the fact that the primary system would be represented in the secondary system by a kind of mirror process.

- And it would be realized also by the fact that primary and secondary processes could induce some marks or attractors in the interpretation system which could then generate intentions, choices, and so on.

- From the viewpoint of computer science, Cardon's system would be an adaptative multi-agent system limited by a network of new kind of agents controlling secondary associative processes and which adapts its behaviour to the environment by operational closure.

- Is it sufficient to get consciousness?

An artificial consciousness is it possible?

- Does Cardon's system answer the question?

- In fact,we must see, at first, that it is not an actual system but just, still now, a working process.

- So, the question is :if such a system were realized, would it pass the Turing's test? The answer is not sure.

- Moreover, we may also ask two other questions :

- First, is it plausible to intend building a « conscious system » while we do not know exactly what conscious-ness is in a living being?

- Secondly, is it plausible to pretend building a thinking system when we do not know exactly how something makes sense in the language?

Conclusion

- As we have no answer to these questions, we must be cautious, if not sceptic, in front of Cardon's theory.

- It is certainly important that biology and computer science ask questions about consciousness and try to know what it is made of.

- But, as consciousness has degrees, maybe it would be better to study the lowest levels of it before tackling the big question of human consciousness.

- Moreover, there is today, actually, no consensus in social sciences about the facts and properties upon which an artificial conscious system should be grounded.

- So an artificial conscious system is today, and stands a good chance of remaining for a long time, if not for ever, an utopy.

Bibliography

- Cardon, A. , *Conscience artificielle, systèmes adaptatifs*, Paris, Eyrolles, 2000.

- Dennett, D.C., *The intentional stance*, LIT Press, 1987.

- Descartes, R., *Œuvres*, Paris, PUF, 1953.

- Dretske, F., *Explaining behavior*, Cambridge, MIT Press, 1988.

- Edelman, G.M., *Biologie de la conscience*, Paris,O.Jacob, 1992.

- Hofstadter D., *Gödel,Escher,Bach, an eternel golden braid*, New York, Basic Books Inc Publishers, 1979.

- Husserl, E., *Méditations cartésiennes*, tr.fr.,Paris, Vrin,1966.

- Kant, E., *Critique de la Raison Pure*, tr.fr. Paris, PUF,1972.

- Missa, J.N., *L'esprit-cerveau*, Paris, Vrin, 1993.

- Searle, J.R., *La redécouverte de l'esprit*, tr.fr. Paris, Gallimard, 1992.

- Varela, F., *Autonomie et connaissance*, Paris, Seuil, 1989.