# Elephants do not play chess

In Elephants do not play chess, Rodney Brooks claims that classical AI (GOFAI) is basically flawed because it is based on the symbolic system hypothesis. This is why he proposes an alternative way of creating AI agents which he calls Nouvelle AI.

He further claims that in Classical AI, intelligence is decomposed into several modules which taken separately do not perform any behaviour. Thus, the performance of behaviour is achieved by the modules only when taken as a whole. Consequently, in order to improve system behaviour it is needed to improve the single modules.

On the other hand, he devises Nouvelle AI in a different way. Indeed, while intelligence is decomposed into modules as in Classical AI, every module provides a certain behaviour. Thus, in order to achieve more complex behaviour it is possible to combine already existing modules. When the system lacks some basic behaviour, it is possible to add a module in the system.

As we can see, Nouvelle AI is based on a modularisation system which in a certain sense is based on current principle of programming as complex functions are based on an aggregation of simpler functions.

## The symbol system hypothesis

Classical AI was based on the symbol system hypothesis which claims that manipulating a set of symbols is sufficient to generate intelligence (John Searle’s experiment contrasts this view and the qualification problem too).

Thus, according to the symbol system hypothesis, it is possible for a reasoner agent to manipulate a bunch of symbols (which are meaningless to it) and still be intelligent.

In the symbols system hypothesis, the world is described through logical sentences (usually first order logic) and the agent can infer new statements through an inference process.

Therefore, symbols represent entities in the world. They may be individual objects, properties, concepts, desires, emotions, nations, colours, libraries, or molecules, but they are necessarily named entities.

What is remarkable in Classical AI is that while symbols are meaningless to the agents, they are meaningful to the programmers.

A negative consequence of the symbol system hypothesis is that the symbols must be task dependent. Indeed, it is not possible to devise a general purpose symbolic representation of the world, but programmers must focus on a given task in order to create a consistent symbolic representation. The lack of the symbolic system hypothesis to be able to represent the whole world around us makes it impossible to create the so-called General purpose artificial intelligence.

Another negative aspect of the symbol system hypothesis is that in order to accurately represent the real word, objective true statements must be inserted in the world’s symbolic representation of agents. This leads to cumbersome programs and furthermore we humans do not have an objective truth for every aspect of the world around us. This is why probabilistic reasoning systems can deal with real world problems in a more consistent way than symbolic ones.

The **frame problem** is another problem linked to the symbol system hypothesis.

The frame problem consists of the fact that after an agent performs an actions in its environment, it must know what has been affected by its actions. In other words, he must know which of its symbolic representation have changed and which ones have not changed:

Imagine being the designer of a robot that has to carry out an everyday task, such as making a cup of tea (with explicitly stored, sentence-like representations of the world).

Now, suppose the robot has to take a tea-cup from the cupboard. The present location of the cup is represented as a sentence in its database of facts alongside those representing innumerable other features of the ongoing situation, such as the ambient temperature, the configuration of its arms, the current date, the colour of the tea-pot, and so on. Having grasped the cup and withdrawn it from the cupboard, the robot needs to update this database. The location of the cup has clearly changed, so that's one fact that demands revision. But which other sentences require modification? The ambient temperature is unaffected. The location of the tea-pot is unaffected. But if it so happens that a spoon was resting in the cup, then the spoon's new location, inherited from its container, must also be updated.

How could the robot limit the scope of the propositions it must reconsider in the light of its actions? In a sufficiently simple robot, this doesn't seem like much of a problem. Surely the robot can simply examine its entire database of propositions one-by-one and work out which require modification. But if we imagine that our robot has near human-level intelligence, and is therefore burdened with an enormous database of facts to examine every time it so much as spins a motor, such a strategy starts to look computationally intractable.

As we can see, the frame problem is a real issue as if one agent really has a human-like intelligence then it should check while it is performing an action which sentences its knowledge base change and which ones do not. Checking all sentences in a human-like symbolic knowledge base is an intractable computation.

## The physical grounding hypothesis

Nouvelle AI is based on the physical grounding hypothesis. The physical grounding hypothesis is a total contraposition to the symbolic system hypothesis. Indeed, according to the physical grounding hypothesis the best model of the world is the world itself, thus, it is unnecessary to create a layer of symbols to represent it. The key observation is that the world is its own best model. It is always exactly up to date. It always contains every detail there is to be known.

In order to adopt the physical grounding hypothesis for building agents, they must be capable of sensing the world and interact with it. Thus, agents must not be seen as only functions but as architecture as well.

As we can see the physical grounding hypothesis resembles the Embodiment concepts.

A consequence of the physical grounding hypothesis is the following:

The constructed system eventually has to express all its goals and desires as physical action, and must extract all its knowledge from physical sensors.

(The concepts explained by Brooks can be easily explained by the fact that he is an expert in robotics. As such he could be biased in expressing these concepts.)

Brooks supports the physical grounding hypothesis through the theory of evolution. Indeed, we and animals in general are the way we are now through a long process of evolution that has started millions of years ago. Through the interaction with the dynamic environment and the sensing of our surroundings we have been able to develop our intelligence. Thus, it may be the case that the same holds for machines given that they are physically grounded in the environment as well.