Meaning Engines — Revisiting the Chinese Room

Rodrigo Ventura and **Carlos Pinto-Ferreira**
Instituto de Sistemas e Robótica
Instituto Superior Técnico
Av Rovisco Pais, 1
1096 Lisboa Codex
Portugal
Tel. +351-1-8418271
{yoda,cpf}@isr.ist.utl.pt

Abstract

In this paper, an approach to the concept of meaning is provided and presented. In particular, the Chinese Room scenario, engendered by John Searle (1980) (Searle 1980) is revisited through the introduction of an apparently minor detail in the set-up. From this modification, an architecture that is claimed to provide meaning to objects and situations is presented. This architecture, based on the idea that stimuli should be processed under two different perspectives — a cognitive and a perceptual one — underlies what is here defined as emotion-based agents. As these agents are capable of generating new associations, it is claimed that this learning of meaning mechanism is a meaning engine.

"A topic of world-shaking importance, yet dealt with facetiously; an android trait, possibly, he thought. No emotional awareness, no feeling-sense of the actual meaning *of what she said. Only the hollow, formal, intellectual definitions of the separate terms."*

"Do Androids Dream of Electric Sheep?,"
Philip K. Dick (Dick 1996)

Human beings are remarkably good at assigning meaning to things and situations. Even when faced with abstract pictures, fuzzy clouds, or drifting smoke, people are almost always capable of "seeing" things, faces, animals, or extracting "meaningful" signs. Furthermore, they are not only interpreters of stimuli, but can also generate meaning through a very peculiar process of chaining associations: from a cloud to a face, from a face to bad fate, and so on. Basic, rudimentary learning seems to be rooted in this process of transforming meaningless to meaningful. Also, certain superstitious behaviors (which can be observed even in animals under controlled experimental conditions) depend on the assignment of meaning to experienced situations.

On the other hand, computer programs — even the most sophisticate ones — are syntactical interpreters, parsing sentences and inferring conclusions which are correct provided that input data be true in the modeled world and that the inferring mechanism be sound.

The Chinese room, introduced by John Searle (Searle 1980) illustrates this discussion and sheds light on the issue of understanding the differences between human and computer interpreters.

Following the approach suggested by John McCarthy in his essay "Ascribing Mental Qualities to Machines" (1979), it can be considered, at least at first sight, that any agent, making the "right" decision when faced with a certain situation, is assigning meaning to it. For example, following this tentative definition, to an automatic heating system, low temperature "means" cold as it makes the right decision of switching the power on in this situation. Of course, something is missing in this inadequate definition. Taking an incremental approach to the problem of defining what an agent capable of assigning meaning should contain, it can be asserted that, at least, it should be capable of generating new meanings (*i.e.,* establishing new associations). Of course, the mentioned heating system is unable to perform such a task.

However, this process of associating things must be based on pre-existing associations, on top of which brand new meanings are established. But, in the way down to the foundations of such systems, the analysis ought to end up at some point, in basic, essential associations which have to be built-in, hard-wired in the agent. The conclusion that underlying what is usually called as meaning there is in fact a set of primitive associations is crucial to the establishment of systems capable of assigning meaning. Another interesting aspect to be taken into consideration is that these basic associations are species-dependent: as a matter of fact, the color red means different things to bulls and elephants, supposedly because their needs (in terms of survival and adequate decision making) are also different.

Now, a new tentative definition of an agent, capable of assigning meaning, could be tried (see figure 1). What is here defined as a *meaning engine* should include:

1. a (possibly small) number of built-in associations between certain characteristics present in some stimuli and leading to adequate behavior in face of them;
2. the capability of building new associations dynamically, on top of existing ones (recursively learnt or built-in);

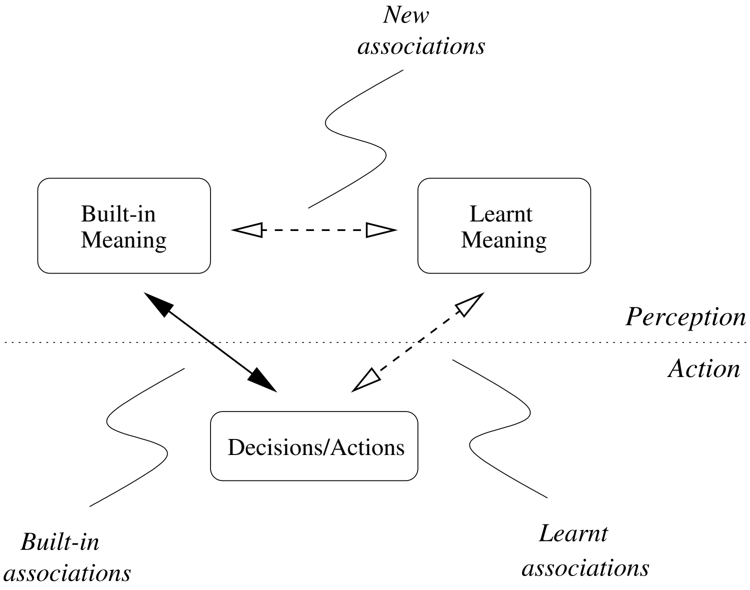

Figure 1: General view of the meaning assignment process.

3. the ability of interpreting both basic and learned associations, so making the "right" decision.

In order to make an effort to implement these ideas, the concept of association should be clarified and given an operational understanding.

To cope with the variety, complexity, and constant modification present in real world stimuli, it is not possible to associate images of these stimuli directly to internal representations or courses of action. Recall that mapping arbitrarily complex images of stimuli to courses of action is an intractable task. Thus, an intermediate compact "meaning engine" is required. A possible way to circumvent this problem is through the simultaneous processing of stimuli under two different perspectives: a cognitive and a perceptual one. The cognitive processor — intended for recognition and reasoning purposes — analyzes the stimuli in terms of finding known patterns. On the other hand, the perceptual processor extracts significant features of the stimuli, for instance, object dimensions, optical flow, color, bright intensity, and so on. This pair of processors provides two "images:" a cognitive, and a perceptual (Ventura & Pinto-Ferreira 1998; Ventura, Custódio, & Pinto-Ferreira 1998).

For example, when faced with the image of a moving object, the cognitive processor provides elements to recognition (is it a lion or a rabbit?) whereas the perceptual processor delivers an assessment of the prevailing color, moving speed, dimension, and other relevant characteristics found in the scene (is it a huge object with a particular color — a predator, or a little quick moving object — a prey?). These characteristics compose a "perceptual image" that serves two purposes: on the one hand, it allows a rough evaluation of the situation for decision making (run away or prepare to attack?) and, on the other, it helps the search which underlies the process of recognition: instead of comparing the "cognitive image" under processing with all elements stored in memory, the search is bound to those objects sharing the same perceptual image. To reach this objective, cognitive images and perceptual ones are associated and memorized in such a way that the later indexes the former.

Systems incorporating this double processing and knowledge representation mechanism, indexing and storing two representations of the same object are defined as emotion-based agents[1].

The association cognitive/perceptual with respect to a stimulus provides meaning to it. The above mentioned double knowledge representation mechanism can now be exposed to the modified Chinese room scenario. Imagine that the slip containing Chinese characters includes some of them colored with the following built-in associations accepted by you, the reader, who is in the room: red — danger; pink — love; blue — friendship; yellow — hunger; green — food; and black — no *a priori* connotation.

According to the traditional Chinese Room setup, assume that inference is performed processing characters irrespective of their color. In the modified version introduced, a single colored symbol may propagate its meaning throughtout the chaining of reasoning, coloring previously uncolored symbols, *i.e.,* assinging meaning to previously meaningless characters.

Can you learn and understand Chinese in such a scenario? As you receive news about the outside world and since you are capable of inferring sound conclusions using strings of Chinese characters and rules of inference, you start associating meaning to the inferred characters (even the ones initially not colored) and soon you will start understanding the Chinese language provided that you are allowed to ask questions (obviously printed in Chinese characters).

It can be objected that colors are, in fact, a kind of intermediate language to help you to translate Chinese. It is true. However, what is here asserted is precisely that the built-in associations taken into account by the perceptual processor provide an essential, preexisting language (common to the human species) on top of which more sophisticate languages can be built. As a conclusion, what was missing in the original Chinese Room was a perceptual processor as it already had a cognitive one.

This ensures that people have a common decoding machine that allows them to interpret each other's messages, at least, at a very basic level.

If the hypothesis above formulated is correct, we can send and receive understandable messages to culturally distant people, in terms of space and time, that is to say, it is possible to communicate even with people in the far future. However, this does not mean that it will be easy or straightforward: as it always happens, to

[1] The reason why emotion-based agents will soon become clear.

understand each other, a mutual effort is necessary.

There are some arguments that support the ideas provided in this paper. On the one hand, there is a strong evidence from neurobiology suggesting the mentioned double processing and knowledge representation mechanism, namely the Cannon-Bard theory ((LeDoux 1996), pg. 82–85) and the Papez circuit theory ((LeDoux 1996), pg. 87–90). Furthermore, recent results from neuroscience indicate that perceptual representations are essential for reasoning, without which subjects become unable to take the simplest decisions (Damasio 1994). On the other hand, an implementation described in (Ventura & Pinto-Ferreira 1998) exhibited some interesting results, namely the ability to learn cognitive representations based on the presented perceptual meaning scheme, and to adapt through time to changes in the environment.

Note finally that if some stimuli become meaningful while others do not, it can be said that a relevance mechanism is implemented. An external stimuli is relevant according to the strength of the meaning assigned to it. And thus this helps the agent to distinguish what in an environment is relevant, among the complexity of stimuli it can provide.

The study and implementation of artificial emotions is a new born field in Artificial Intelligence (Reilly & Bates 1992; Minsky 1988; Picard 1995; Sloman & Croucher 1981; Velásquez 1997; Ventura & Pinto-Ferreira 1998; Cañamero 1997). As Marvin Minsky stated in his "Society of Mind" (Minsky 1988): "The question is not whether intelligent machines can have any emotions, but whether machines can be intelligent without emotions."

References

Cañamero, D. 1997. Modelling motivations and emotions as a basis for intelligent behavior. In *Proceedings of Agents'97*. ACM.

Damasio, A. R. 1994. *Descartes' Error: Emotion, Reason and the Human Brain*. Picador.

Dick, P. K. 1996. *Do Androids Dream of Electric Sheep?* Harper-Collins.

LeDoux, J. 1996. *The Emotional Brain*. Simon & Schuster.

Minsky, M. 1988. *The Society of Mind*. Touchstone.

Picard, R. W. 1995. Affective computing. Technical Report 321, M.I.T. Media Laboratory; Perceptual Computing Section.

Reilly, W. S., and Bates, J. 1992. Building emotional agents. Technical Report CMU-CS-92-143, CMU, School of Computer Science, Carnegie Mellon University.

Searle, J. R. 1980. Minds, brains, and programs. *Behavioral and Brain Sciences* 3:417–457.

Sloman, A., and Croucher, M. 1981. Why robots will have emotions. In *Proceedings IJCAI 1981*.

Velásquez, J. D. 1997. Modeling emotions and other motivations in synthetic agents. In *Proceedings AAAI-97*, 10–15. AAAI.

Ventura, R., and Pinto-Ferreira, C. 1998. Emotion-based agents. In *Proceedings AAAI-98*. AAAI. (accepted for publication).

Ventura, R.; Custódio, L.; and Pinto-Ferreira, C. 1998. Artificial emotions — goodbye mr. spock! (accepted for publication in IBERAMIA-98).