Emotion-based Agents*

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

To survive in a dynamic and rich environment, human beings have to process very complex stimuli in real time. Whatever artificial system satisfying the challenge of achieving a similar performance in complex data handling, ought to incorporate mechanisms specifically conceived to perform efficiently.

We hypothesize that such efficiency oriented systems process stimuli — simultaneously — under two different perspectives: a *cognitive*, elaborative — which allows them to understand what is happening and what they know about the world, and a *perceptual*, immediate — which permits them to react quickly and decide adequately in circumstances demanding urgent action. Hence, from the very same complex stimulus, two sets of facets are extracted: one, mostly directed to recognition and reasoning purposes, and another, aiming at assigning degrees of threat, danger, pleasure, and so on, to the current situation, constructing what we call a *vector of desirability*.

For instance, 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 features 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" which serves two purposes: on the one hand, it allows a rough evaluation of the situation and the corresponding decision making. On the other, it helps the search which underlies the process of recognition: instead of comparing the "cognitive image" under processing with all the elements stored in memory, the search is bound to those objects sharing the same perceptual image. To reach this desideratum, "cognitive images" and "perceptual images" extracted from the same source stimulus should be associated and memorized in such a way that the latter indexes the former.

This kind of system should be bootstrapped by the incorporation of built-in associations. In fact, there should exist some stimuli which are *essential*, innate: for instance, animals faced with their preys or predators decide either to attack or run away as a function of the vector of desirability a perceptual image suggests. This assignment depends on the considered species.

When an unknown object appears in the scene, its perceptual image is extracted and a first estimate of the vector of desirability, as well as the corresponding decision making in the short term is performed. On the other hand, as the agent does not find the corresponding cognitive image in memory (because the object is unknown), a new cognitive-perceptual association is established. This seems to be the mechanism underlying the pavlovian reflex — as it lies on top of built-in associations. This also seems to be the basic mechanism responsible for the assignment of meaning to objects.

Systems incorporating this double processing and knowledge representation mechanism, indexing and storing two representations of the same object together with a vector of desirability, are here defined as emotion-based agents.

This model is corroborated by the work of several researchers in the field of neuroscience. Namely, in Papez circuit theory the functions performed by the hypothalamus and cingulate cortex (in the human brain) can be identified with the cognitive and perceptual processing discussed above (LeDoux 1996). And according to Damasio's ideas, emotions are essential to human decision making (Damasio 1994).

Several architectures have been proposed to deal with real time processing and decision making; in all of them it is possible to identify "cognitive" and "perceptual" levels. However these levels work independently and, in certain cases, one blocks the other. What makes this proposal different and appealing is the intertwining of both, and the consequent learning capabilities which it allows.

An architecture based on these concepts was implemented and the results are encouraging: not only does the agent learn new associations, but it also recognizes objects efficiently and decides correctly when faced with new, unknown situations. And interestingly, what can be considered as an emotional behavior, was observed.

We are afraid that, to reach intelligent behavior, we will end up implementing some of the (apparently) most stupid things human beings exhibit, including bad temper.

References

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

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