Emotion-Based Control Systems
(Preliminary Report)

Rodrigo Ventura Carlos Pinto-Ferreira
Instituto de Sistemas e Robótica
Instituto Superior Técnico
Rua Rovisco Pais, 1
1049-001 Lisboa
PORTUGAL
emails: {yoda,cpf}@isr.ist.utl.pt

According to the model developed by Antonio Damasio [1], the process of decision making depends on the mechanism of emotions. In fact, the handling of unpredictable, dynamic and aggressive environments under partial ignorance, presupposes a highly effective mechanism of situation assessment and decision making. Of course, such a mechanism cannot be rooted solely on deductive reasoning (the one underlying the verbal knowledge representation paradigm) as it is extremely time-consuming.

Beyond the fact that, in the context of Artificial Intelligence (AI), deciding what is the relevant knowledge to reason about is an ill-solved problem, the very process of verbal inference (either chaining or resolution based) is itself too complex and therefore inadequate when dealing with demanding (urgent) situations. This is the reason why the so called knowledge-based control has only been either directly applied to very narrow-bandwidth systems or in the supervision of low level controllers (by changing set points, PID parameters, and so on).

On the other hand, children aged two or less (that is to say, in a pre-linguistic phase) are capable of solving planning problems and make adequate decisions supposedly without using verbal reasoning (this is the so called sensory-motor intelligence referred by Piaget [2]). Furthermore, animals such as cats and dogs exhibit a performance on decision making that would make blush of shame our most competent robots (of course, if they could feel)...

It can be asserted that children aged two or less and such animals rely on reactive planning (the process introduced by Rodney Brooks in the 80's [3]). Notwithstanding being a very interesting and effective mechanism of decision making, the reactive planning paradigm has shown severe limitations whenever we abandon the world of robotic insects and try to implement it in a little bit more sophisticated application (say, in the context of robotic soccer).

Reactions — in human beings, for instance — correspond to responses of the body to stimuli demanding immediate action and they do not presuppose any kind of assessment of the situation, so being delivered without the intervention of the brain [4, 1]. Therefore, the gap between purely reactive systems and inference based agents is so big that it is difficult to conceive the integration of these two paradigms to construct an intelligent and effective agent.

That is to say, there exists a missing link that should handle the assessment of the situation and to decide either to reason or to respond, according to the urgency of the situation [5].

The way Nature has implemented this intermediate level of competence in decision making seems to work as follows. When exposed to a stimulus, an agent performs, simultaneously, two kinds of processing efforts [6]: one, that we call cognitive, which aims at finding out what the stimulus is (being so devoted to recognition purposes) and another, that we call perceptual, which intends to assess whether the situation demands urgent action (being so dedicated to extract features from the incoming stimulus). Of course these two processes differ on sophistication and efficiency: recognition, as it relies on pattern matching is a very heavy process in terms of computation, whereas feature extraction can be very light in the same terms.

Albeit being conceptually different processes (recognition and feature extraction), they should be interrelated in such a way that the later helps the former.

However, as feature extraction ends first than recognition, the clues provided by it constitute a precious instrument to anticipate proper action. Recall that feature extraction can rely on color evaluation, optical flow reckoning, assessment of dimension of objects, determination of sound intensity, and so on.

To illustrate the mentioned mechanism, consider a bull exposed to a moving red object. From the point of view of the bull survival it is much more relevant to finding out that there is something red moving (feature extraction) than to discover that it is say, a handkerchief or a prey (recognition). And this relevance (from the point of view of the bull survival) derives from the fact that red moving objects should be attacked by the bull. This is because evolution installed in each animal, a mechanism to make decisions according to the needs of the corresponding species — the determination of what can be their preys and predators. Notice that the perceptual assessment of a scene (in terms of the relevant features) depends on the considered species and it is a built-in characteristic of it. The association between certain features and corresponding perceptual responses seems to be the basic mechanism of meaning [7].

Suppose now that the situation is such that it does not demand urgent action (according to the result of feature extraction). In that case, there is enough time to explore the stimulus and to end the process of cognitive assessment, having as a consequence, the recognition of what is in the scene.

The intertwining of these two processing efforts — cognitive and perceptual — constitutes the roots of what we call emotion-based agents [8]. The sort of responses got from mammals — for instance — is here called "emotional" because, in common-sense terms, we ascribe emotional qualities to bulls, dogs, and cats (of course, we do not do the same to ants and bees, for instance, probably because these insects rely on a purely reactive approach to decision making). Following the Antonio Damasio paradigm, there is a clear distinction between emotions and feelings. In a very loose way, we can say that feelings correspond to consciousness of emotions, so we hypothesize that only human being are capable of having feelings.

Moreover, the efficiency got from the simultaneous processing of stimuli (cognitive and perceptual) does not result solely from having a system for deciding what to do before recognizing what the agent is facing. There is a much more interesting mechanism based on associations of cognitive images with perceptual images. This mechanism is the one described by Antonio Damasio and is designated as somatic marker [1]. The marking of cognitive images by their perceptual counterparts constitutes a remarkable system of categorization: as groups of cognitive images are indexed by the same perceptual image, whenever a perceptual image is computed the search effort to perform the pattern matching is drastically reduced (as the number of comparisons in memory decreases in a corresponding manner).

One of the interesting aspects deriving from the mentioned architecture is the way such systems handle the problem of assessing the relevance of a certain situation. If the impression got from the perceptual processing is low (that is to say, if the feature extraction for a particular species delivers characteristics which are not "meaningful"), then the situation is not relevant, not deserving special consideration.

To test the presented model, three preliminary implementations were developed [9]. The first one implements a marking mechanism between the a cognitive and a perceptual input. In this implementation the stimulus is divided in a cognitive part and a perceptual one. The agent learns to associate pairs of cognitive and perceptual images, so that when exposed with similar cognitive images, it recalls the perceptual ones, and decides (a simple yes/no decision) according to past experience. This process is not rigid, in the sense that the implementation provides a set of parameters that can be adjusted in order to simulate various kinds of behaviors, such as superstition and stubbornness (loosely speaking).

In the second implementation, the agent is faced with a sequence of 16x16 pixel images. To bootstrap the system, green pixels are associated with a "positive" stimulus, while red pixels are "negative." The agent associates shape (cognitive image) with this positiveness/negativeness (perceptual image), so that when it is a posteriori faced with black and white images, it classifies them according to those previously established associations. Furthermore, it is able to override the basic perceptual color-based assessment with the mark from a more refined match at the cognitive layer.

Finally, a third implementation attempts to reproduce the experiment Damasio applied to both frontal patients (brain lesions in the prefrontal cortex) and normal controls [1]. This experiment consists on four deck (named A through C) from which the player draw cards. In each turn, the player has to decide from what deck does he want to draw. After choosing a deck to draw a card, the player is informed of an amount of money earned/lost with that card. Initially he starts with, say, $2000. In a simplified version of the original game [10, 11], decks A and B usually give $100 except for a few cards that make the player lose -$1250, while decks C and D usually give a lower value of $50 where there are more frequent losses of -$250. The net profit of decks A and B is negative, while decks C and D provide a positive net profit. In Damasio results, normal patients tended to choose decks C and D, since they provide a long term profit, while frontal patients preferred the other two decks. According to Damasio, frontal patients lacked the brain mechanism responsible for associating emotional assessments to situations. The results obtained by the implementation appear to corroborate these two distinct behaviors, i.e., the decks chosen by the full implementation prefers decks C and D, while a modified version of it (stopped from recalling past associations) prefers decks A and B.

The relevance of the present model to the control and the supervision of systems lies on the fact that, in this context, it is very important to respond quick and efficiently to unexpected situations, by learning associations between current situations and control strategies. The inputs and the state variables of a system can be considered as stimuli to feed a double processing system as described above. The cognitive image can be considered as the set of values collected in a time frame. On the other hand, the perceptual image can result from the determination of certain characteristics (taken separately of in combination) as overshoot, rate of variation of state variables, and so on. The next step is to establish a basic set of associations in order to allow the system to respond to urgent situations (solely based on the perceptual image). As the supervisor starts marking cognitive images with perceptual ones (a basic mechanism of learning), it becomes able to anticipate those situations (this is what humans apparently do when using the somatic marker). On the other hand, the matching of a certain configuration with one previously stored in memory can be assessed in terms of the positiveness or negativeness of the present situation by consulting the cognitive/perceptual mark.

The control and supervision of large scale, non-linear, and non time-invariant systems ought to incorporate planning and decision making mechanisms together with low-level controllers, integrated in such a way that performance (both in terms of learning, quality of response, and efficiency) is ensured.

References

[1] Antonio R. Damásio, *Descartes' Error: Emotion, Reason and the Human Brain*, Picador, 1994.

[2] Jean Piaget and Bärbel Inhelder, *The Psychology of the Child*, Basic Books, Inc., 1969.

[3] Rodney A. Brooks, "A robust layered control system for a mobile robot," *IEEE Journ. of Robotics and Automation*, vol. RA-2, no. 1, pp. 14–23, March 1989.

[4] Joseph LeDoux, *The Emotional Brain*, Simon & Schuster, 1996.

[5] Rodrigo Ventura, Luis Custódio, and Carlos Pinto-Ferreira, "Emotions — the missing link?," in *Emotional and Intelligent: The Tangled Knot of Cognition*, Dolores Cañamero, Ed., 1998, pp. 170–175.

[6] Rodrigo Ventura, Luis Custódio, and Carlos Pinto-Ferreira, "Artificial emotions — goodbye mr. spock!," in *Progress in Artificial Inteligence, Proceedings of IBERAMIA'98, Lisbon, Portugal.* 1998, pp. 395–402, Colibri.

[7] Rodrigo Ventura and Carlos Pinto-Ferreira, "Meaning engines — revisiting the chinese room," in *Workshop: Grounding Emotions in Adaptive Systems*, Dolores Cañamero, Chisato Numaoka, and Paolo Petta, Eds. SAB'98: From Animals to Animats, August 1998, pp. 68–70.

[8] Rodrigo Ventura and Carlos Pinto-Ferreira, "Emotion-based agents," in *Proceedings AAAI-98.* AAAI, 1998, p. 1204, AAAI Press and The MIT Press.

[9] Rodrigo Ventura and Carlos Pinto-Ferreira, "Emotion-based agents: Three approaches to implementation (preliminary report)," in *Workshop on Emotion-Based Agent Architectures (EBAA'99)*, Juan Velásquez, Ed., May 1999, pp. 121–129.

[10] Antoine Bechara, Antonio R. Damásio, Hanna Damásio, and Steven W. Anderson, "Insensitivity to future consequences following damage to human prefrontal cortex," *Cognition*, vol. 50, pp. 7–15, 1994.

[11] Antoine Bechara, Hanna Damásio, Daniel Tranel, and Antonio R. Damásio, "Decising advantageously before knowing the advantageous strategy," *Science*, vol. 275, pp. 1293–1295, February 1997.