THE AUTONOMOUS SYSTEMS LABORATORY
UNIVERSIDAD POLITÉCNICA DE MADRID

Emotion in the ASys Framework

Ricardo Sanz, Ignacio López,
Carlos Hernández, Jaime Gómez

ASLab A-2008-003 v 0.0 Draft
February 15, 2008

The construction of fully effective systems seems to pass through the proper exploitation of goal-centric self-evaluative capabilities that let the system teleologically self-manage. Emotions seem to play this kind of role in biological systems and hence the interest of emotion for artificial machines. Beyond the obvious capability for displaying human-like emotion-ladden faces in robots, emotional mechanisms may play a critical role in the sustainment of function in changing *-e.g.* interactive-environments. This paper describes the role that emotion is intended to play in the ASys Framework for conscious autonomy.

1 Introduction

The central tenet of artificial systems research is the development of engineering capability for achieving some desired level of performance in a technical system: 220 volt in a wall socket, 240 k/m in a car, 80 Hz beat in a pacemaker, *etc.* .

Once the development of the base technology —electrical engineering, mechanical engineering, embedded electronics— lets achieve this performance level a second aspect gains in importance: maintaining this performance. This is obvious in the pacemaker or wall socket; it is less obvious in the car but just for the speed (*e.g.* no for the brakes).

1.1 Sustainable performance

The preservation of performance levels shall be understood in two different but equally important aspects: i) as *sustainability* and ii) as *resilience*. Using some more words we can state the core maxima of engineering today as *build systems to keep the adequate performance level doing a sustainable impact on their environment.* This idea of sustaining performance obviously maps in different ways to different kinds of applications that have more or less resilient structures in changing environments. In our research of *robust autonomy* we focus on two domains of artificial systems: large-scale industrial plants and mobile robotics.

To achieve this capability of *robust autonomy*, the material and energy flows from/to the environment ought to be accompanied by the associated flows of information that enable the systems to manage the uncertainty in the operational conditions. At the end, the system uses this information to adapt itself to the changing circumstances so as to maximise its effectivity. It seems that the construction of maximally effective systems seems to pass through the proper exploitation of goal-centric self-evaluative capabilities that let the system teleologically self-manage.

1.2 The ICEA Project

In the context of the EU-funded ICEA Project[1] this search for resilience is focused on the way that cognitive, emotional and autonomic subsystems are integrated into a single control architecture: the mammal brain.

The special focus on emotions is due to the fact that emotions —as internal states and processes— seem to play this kind of role of valence-centric modification in biological systems. Several brain subsystems are being investigated in this direction (especially basal ganglia, amygdala and hippocampus).

Beyond the obvious capability for displaying human-like emotion-ladden faces in robots (see Figure 3), emotional mechanisms may play a critical role in the sustainment of function in changing -*e.g.* interactive- environments.

This paper describes the role that emotion is intended to play in the ASys Framework for conscious autonomy.

2 Models of Emotion

A common, somewhat folk, theory of human emotions say that they are, to a large extent, subjective and non deterministic. This accounts for the obvious fact that apparently identical stimuli may raise different emotions in different humans, and the same individual may experiment different emotions in response to the similar

[1] www.iceaproject.eu

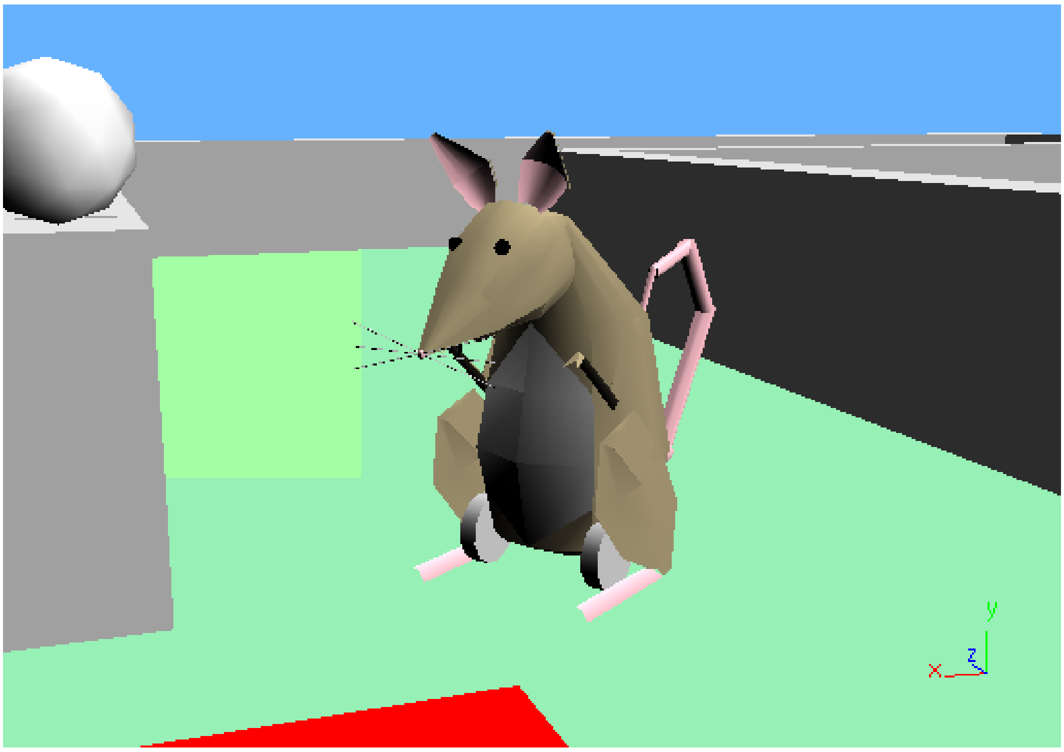

Figure 1: The IST ICEA project is focused in the extraction of integration patterns among the cognitive, emotional and autonomic systems of the rat. These are evaluated on technical systems including physical and simulated rats.

stimulus. Obviously, and from a purely systemic perspective, there must be some disparity between systems if the behaviour differs but the ascription of the variety in emotional response may go from subtle details in otherwise apparently identical stimulus to the very possibility of human freedom and non-determinism.

2.1 Understanding Emotion

The usual understanding of emotions falls into several related aspects:

A1 how emotional behaviour is *triggered* by event surrounding the agent;

A2 how emotion is *manifested* —displayed— by/in the agent; and

A3 how emotion is *felt* by the agent.

This heterogeneity notwithstanding, it is necessary to think about basic physiological principles going down to neural-hormonal mechanisms that make a particular event 'emotional'.

2.2 Classical models

The classical theories of James-Lange or Cannon-Bard address the causality of the relation between A1-A2-A3. The model of James-Lange states that in animals, the triggering (A1) are experiences in the world, the autonomic nervous system then creates physiological events such as increased heart rate or muscular tension (A2), and then emotions come as conscious feelings which come about as a result of these physiological changes (rather than being their cause as the Cannon-Bard model postulates).

Plenty of models can be found in the literature among which we would like to distinguish —for their closeness to our own model— the models of Arnold (1960) for its relation with the shaping of action tendencies; the model of Frijda (1987) from the perspective of emotions constituting forms of action readiness; or the models of Plutchik and Kellerman (1980) and James (1884) in relation with the bodily basis of adaptive mechanisms.

2.3 Damasio's Model

In this sense, one of the most complete models is that of Damasio's. The somatic marker hypothesis Damasio (1999) and related machinery can be used to provide a deeper understanding of emotional system organisation (see Figure 2). This structure is among the architectures explored in the ICEA project in order to provide a coherent picture of the integration of cognitive, emotional and autonomic aspects.

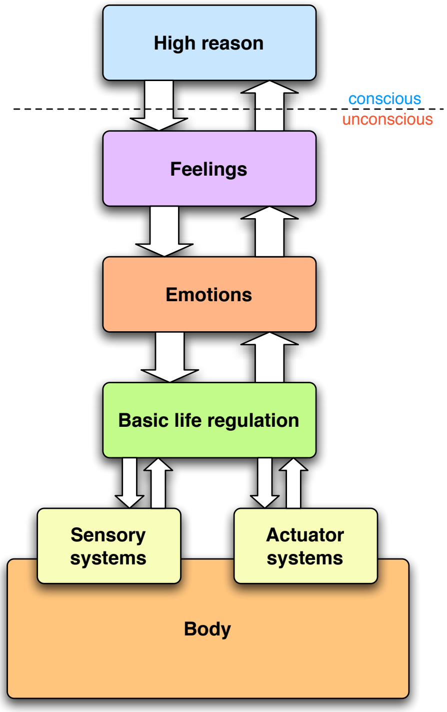

Figure 2: Damasio proposes a hierarchical process for emotion-raising distinguishing between basic autonomic mechanisms, emotion, feeling and conscious awareness of emotion.

Damasio's is a concrete proposal in line with what is needed from a scientific and technical approach to emotion and cognition Ortony et al. (1988). Architectural approaches go beyond the three behavioural and phenomenal aspects mentioned before —triggering, display and feeling— addressing what are the physiological mechanics for all this functioning.

2.4 Assessing Models

Most data concerning emotion comes from experimentation on animals and humans. Data coming from these sources are widely heterogeneous; from single neuron firing patterns, columnar behaviours or activation levels of a whole brain area to verbally reported psychological data.

Some theoretical models of emotion and associated computational implementations are being explored as a promising tool for integrative understanding of emotive-cognitive mechanisms. The main value this approach offers is the possibility of having a precise, more rigourous method to grasp the core concepts and architecture.

In this sense we can cite Botelho (2001):

> We present a preliminary definition and theory of artificial emotion viewed as a sequential process comprising the appraisal of the agent global state, the generation of an emotion-signal, and an emotion-response. This theory distinguishes cognitive from affective appraisal on an architecture-grounded basis. Affective appraisal is performed by the affective component of the architecture; cognitive appraisal is performed by its cognitive component.

Emotions affect all levels of operation in a system, from basic life regulation to conscious, cognitive processes. We use the term *transversal* to indicate this fact. The way in which system operation is affected is specific to each level. Within each level, it is specific to each organ, component and process.

In other words, this means that emotions provide a common *control broadcast* infrastructure which may be used differently by each of the processes in the system. In natural systems, emotions may be conveyed by hormones (or any other bodily broadcasting mechanism). This mechanism must be shared by many organs and processes in the system, which will interpret hormones according to their purposes and architectures. For instance, a cognitive process may interpret emotions to obtain auxiliary information for making a decision regarding what the system must do next. The same hormones may be interpreted concurrently by other processes in order to detect danger, risk, or a need to obtain food, for example.

Artificial implementations of emotions are not developed yet to the same degree as the natural. However, large, distributed, fault-tolerant systems include mechanisms which play a similar role. Fault detection, damage confinement, error recovery and fault treatment are based on broadcast messages and other mechanisms shared and used by system components in analogous ways to the natural counterparts.

This globality and multi-level character of emotions explains some distinctions of emotion-relative phenomena present in the literature, such as Damasio's Damasio (1999, 2004):

- state of emotion,
- state of feeling an emotion,
- state of a feeling of an emotion made conscious.

One way in which emotions are transversal is by broadcasting a summarised picture of the system state to many of its components and processes. This means not only a summary of how is components find themselves, but also a certain sense of affordance of the current scenario relative to the current system situation, processes and objectives.

This is useful to the system in order to adapt to its scenario of operation, mainly for three reasons:

- Emotions are fast, and are available before other more cognitive information.
- Emotions, being to some extent global, contribute to co-ordination and focus of large quantities of system processes and components, which is a factor for preserving system cohesion.
- Emotions can be externalised and hence used for behavioural organisation (cooperation and competition are examples) in multi agent environments (societal behaviour being the clearest example).

This last aspect —that of externalisation of emotional states has rendered emotional expression one of the main topics of emotion research (Figure 3) Darwin (1872); Ekman (1982).

There are plenty of efforts into the implementation of emotion in machines as inspired by biosystems Trapl et al. (2002) but in most of the cases they have a serious neglect of addressing the core issues and just focusing on mimicking shallow, observable manifestations of emotion (*e.g.* making robot faces *a la* Ekman). This work is irrelevant insofar it does not address a sound theory behind it. All that is expected is some improvement in social capability by facial displaying for human emotions. This is hopeless because the functional value of the display of an emotional state in a social interaction is based on the activation in the receptor of the

Figure 3: A big amount of research on emotion has been focused on the expression of facial emotion neglecting the inner functional aspects of it.

display of a behavioural model of the displayer so as to maximise effectiveness of interaction.

Mimicking faces is then useless unless the operational state of the displayer is what is captured in the model going to be activated in the receptor, and clearly this is not the case of human vs. robot architectures (*i.e.* the mental model of the receptor will be a model of a human whereas the robot isn't a human at all from an architectural functional point of view). This issue has been widely addressed in the field of human-computer interaction and the mental models community Gentner and Stevens (1983). What is important then is the raising of mental states in the reception (*i.e.* the activation of mental models) that are relevant for the interaction. This only can be done if emotion is tightly tied to the inner operational mechanisms of the agentConde (2005).

Facial expression of emotion is an externalisation of emotional state to help reconfigure a multi-agent organisation taking into account individual agent operational states. This has been implemented in our framework by means of a mapper from technical operational states into facial displays so as to better understand the technical status of the machine by the human (see `EmotionMapper` in Figure 4).

The topic of emotion, from an engineering point of view, is then reduced to three core aspects:

- What is the general form of an emotional mechanism?
- What is the function that emotional mechanisms do play?

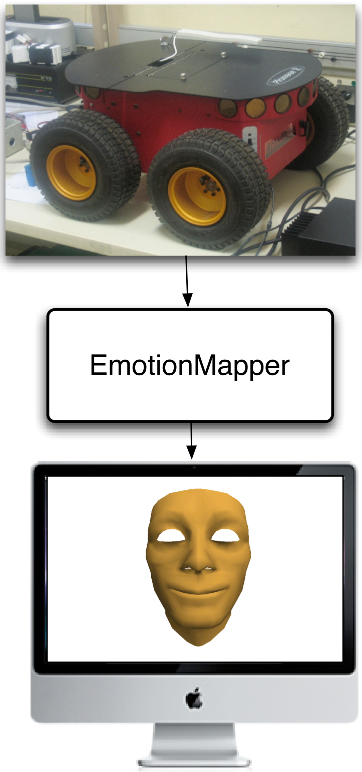

Figure 4: Facial expression of emotion is an externalisation of emotional state to help reconfigure a multi-agent organisation taking into account individual agent operational states.

- What is the best strategy for emotion implementation?

We will analyse these aspects in the context of a technical framework for the engineering of maximally autonomous systems.

3 Self-awareness Models

The mechanisms of emotion impinge on the behavioural capability of the agent so as to prepare it for future action. This makes emotion a core capability for sophisticated self-management control architecture where outer control loops —emotion— determine the functioning of inner control loops —homeostasis— so as to maximise survivability.

Damasio's model on consciousness lays out another control loop atop of these two (see Figure 2) rendering a high-level reasoning capability.

As was the case with emotions, there are plenty of models of consciousness that try to address the relation of physiology and the three core aspects of consciousness: world-awareness, self-awareness and qualia. We can distinguish as maximally relevant for our work —due to their abstract, general nature— the global workspace model of Baars Baars (1997) and the information integration model of Tononi Tononi (2004).

The integrated control model of consciousness Sanz et al. (2007) is based on the provision of self-awareness by means of model-based perceptual mechanics. This model of consciousness is part of a general framework for autonomous systems engineering: The *ASys Framework*.

4 The ASys Framework

The ASLab ASys Project is a long-term research project focused in the development of technology for the construction of autonomous systems. What makes ASys different from other projects in this field is the extremely ambitious objective of addressing *all the domain of autonomy*. We capture this purpose in the motto "engineering any-x autonomous systems". The ASys Framework is both a theoretical framework for understanding all the relevant issues and a technological framework that enables the technically sound creation of autonomous systems (where autonomy is understood in its broadest sense and not in the severely restricted sense of the term *autonomous intelligent systems* that is usually equated to *mobile rbehaviourobots*).

One of the central topics in the ASys Framework is the pervasive model-based approach. A truly autonomous system will be continuously using models to perform its activity. An ASys will be built using models of it. An ASys can exploit its own very models in driving its behaviour. Model-based engineering and model-based behaviour then merge into a single phenomenon: model-based autonomy. We equate this with *cognition*.

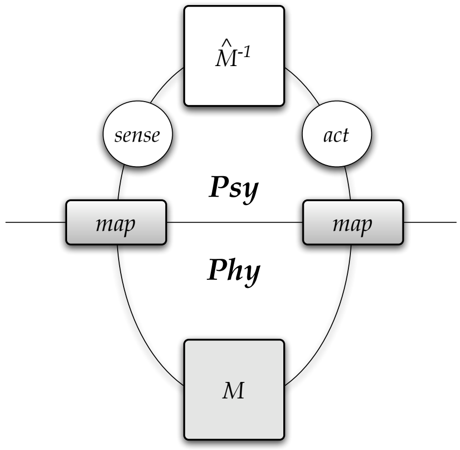

Figure 5: The mind as model-based controller vision is a central concept in the ASys Framework.

The ASys Framework hence establish that a system is said to be cognitive if it exploits models of other systems in their interaction with them. Models and knowledge are then equated and the ASys Framework provides a link between ontological and epistemological aspects of mind.

The ASys Framework follows a principled approach to autonomous system mind construction —the *cognition as model-based behaviour* being the first principle— so as to ground a systematic engineering approach that shall end in rendering machine consciousness Sanz et al. (2007).

These principles establish guidelines for the systematic, formally grounded development of a real-time control software framework based on the control and software principles of the Integrated Control Architecture Sanz et al. (1999). This will render a methodology, a toolset and an execution framework for the engineering of robust autnomous systems based on the implementation of cognitive mechanisms up to the level of consciousness Sanz et al. (2005).

5 Emotion, Transversality and Structural Reconfiguration

5.1 What emotions are and are not

The analysis of the several models of emotion and consciousness produce some conclusions regarding what emotions are not:

- They are not just sophisticated input handling, *i.e. not just reacting to bears*.

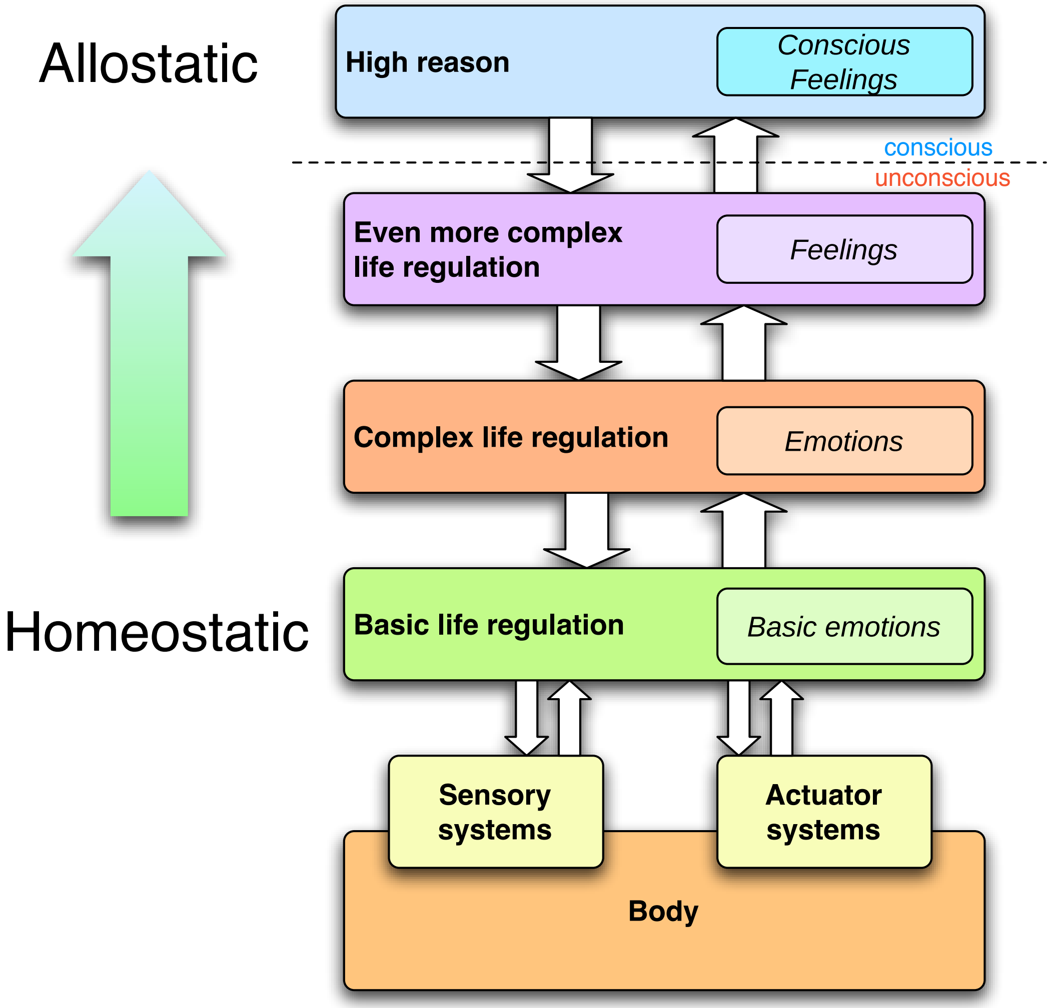

Figure 6: Damasio's layering of emotion appears as labelling of the transversal emotional mechanisms across a layered architecture for control.

- They are not just sophisticated action generation for social affective behaviour, *i.e. not just showing embarrassment*.
- They are not just re-goaling, *i.e. not just change to going to eat when doing sex*.

A deeper analysis abstracting from the biological mechanics into the functional structure renders some conclusions about how emotions work:

- Emotions do generate *synthetic compact states* (performing state space reduction) for the effective tuning and use of evolutionary meta-controllers.
- Emotions do change the *control structures* at the component functions (patterns and roles) of subsystems.
- All this is done in a global controller configuration approach rendering a *transversal structural feedback* architecture.

5.2 Aspectual nature of emotions

The ASys perspective on cognition/emotion goes beyond Damasio's approach of putting emotions/feelings as additional layers in hierarchical controllers. Emotion is no longer another layer in the architecture but a transversal mechanism that crosses across all layers. This is indeed a well known fact in the studies of emotion. Emotions do appear from the subconscious plain to the conscious surface, affecting alll levels in the cognitive structure, from the physiological up to the cognitive, social, self-conscious level.

This implies (see Figure 6) that emotional mechanics are part of each level of the control hierarchy. The level of focus of the analysis is what determines the labelling used for this mechanism: basic emotion, emotion, feeling, conscious feeling, *etc.* In the language of information technology we would say that emotion is an *aspect* Filman et al. (2004) of the different systems that constitute the body and mind of an autonomous agent. From a functional perspective we can also observe that the goals pursued by such control structures go from the purely homeostatic mechanisms for life survival to the higher-level, socially-originated allostatic mechanisms for social behaviour. From hunger to embarrassment, emotions do share the meta-control capabilities over basic behavioural structures.

5.3 Emotion mechanics in the ASys Framework

The ASys Framework for autonomous systems is based on an architecture for software-intensive, distributed, real-time control called the Integrated Control Architecture (ICa) Sanz et al. (1999). This architecture is based on the implementation of patterns of activity across sets of distributed real-time agents. These patterns respond to the needs of the control task that can follow a multilayered, multi-objective control strategy Alarcon et al. (1994).

The implementation of a controller over ICa renders a collection of interacting software components that realise patterns of activity as sequences of service requests. The software component model is of extreme importance in the implementation of such controllers because it provides a common modelling framework both for the physical components of the system under control —the organs in a biological system— that constitute the physical infrastructure and the mental components that constitute the control superstructure. Figure 7 shows three such components in a simple, layered control structure: an organ, a controller and a meta-controller.

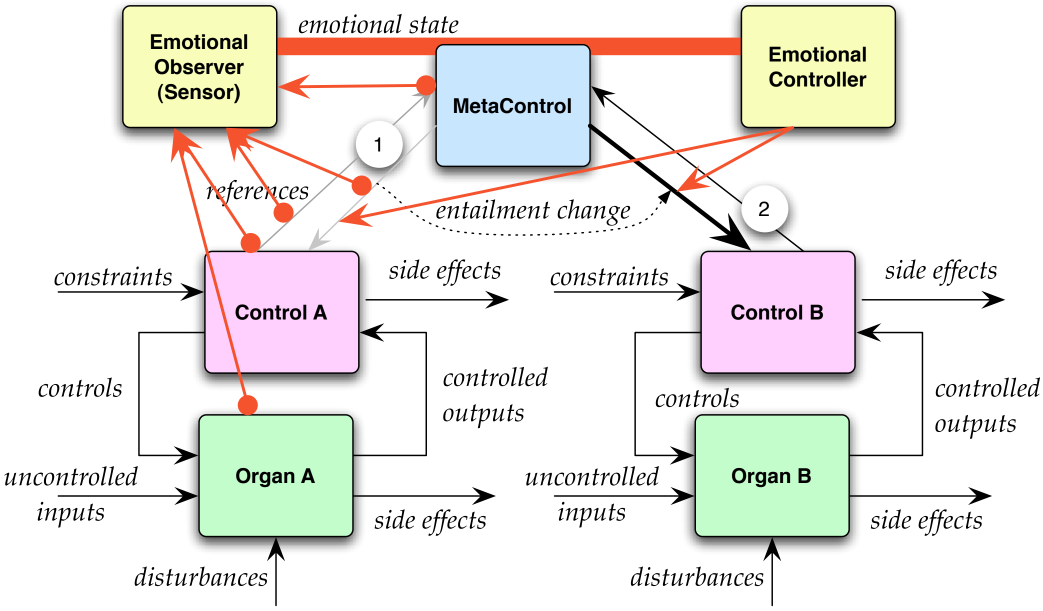

Figure 7: The primary effect of emotion is the change in the functional organisation of the control system of the body. In the figure the emotional system changes the functional organisation from (1) to (2). Both the emotional observation and control are don in terms of a value systems for the agent. This happens in a multi-scale, multilayer organisation that constitutes the integrated global controller of the agent.

Emotions in the ASys Framework, as understood from the previous description, provide the structural mechanisms for control pattern adaptation to the current state of affairs. Examples of this kind of architecture are already available in the world of control systems (for example fault-tolerant controllers and sliding mode controllers).

Now we can provide the ASys Framework answers to the three core aspects of emotion mentioned before:

- What is the general form of an emotional mechanism? A self-reorganising meta-controller.
- What is the function that emotional mechanisms do play? Provide value-centric functional reorganisation.
- What is the best strategy for emotion implementation? Functional modularisation of control functions and integration over a common infrastructure.

6 Conclusions

The paper has reviewed some of the common approaches to emotion understanding, with an special emphasis on Damasio's model of emotion and feeling.

The paper has analysed the extended functional role that emotions can play in complex adaptive controllers and how the different aspects of emotion —triggering, emotional states, bodily effect— are addressed from this perspective.

This understanding has been put in the context of the ASys Framework, a theoretical and technical framework for the implementation of autonomous systems. This framework is based on the construction of modular, component-based control systems following the architectural guidelines of the Integrated Control Architecture (ICa) —a software architecture based on distributed real-time objects.

This framework is being applied to the modelling and understanding of autonomic-emotional-cognitive integration aspects in the rat brain and the implementation of embedded controllers in the context of the IST ICEA project.

7 Acknowledgements

Authors would like to acknowledge the support coming from the European Community's *Seventh Framework Programme FP6/2004-2007* under grant agreement IST 027819 ICEA — *Integrating Cognition, Emotion and Autonomy* and the Spanish *Plan Nacional de I+D* under grant agreement DPI-2006-11798 C3 — *Control Cognitivo Consciente*.

References

Alarcon, I., Rodriguez-Marin, P., Almeida, L., Sanz, R., Fontaine, L., Gomez, P., Alaman, X., Nordin, P., Bejder, H., and de Pablo, E. (1994). Heterogeneous integration architecture for intelligent control systems. *Intelligent Systems Engineering*, 3(3):138–152.

Arnold, M. (1960). *Emotions and Personality*. Cambridge University Press.

Baars, B. J. (1997). In the theatre of consciousness. global workspace theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies*, 4:292–309.

Bermejo Alonso, J. (2006). A state of the art on emotion. Technical Report R-2006-002, UPM Autonomous Systems Laboratory, Universidad Politécnica de Madrid.

Botelho, L. (2001). Machinery for artificial emotions. *Cybernetics and Systems: An International Journal*, 32:465–506.

Conde, R. P. (2005). Mapeo facial de emociones sintéticas. Master's thesis, Universidad Politécnica de Madrid.

Damasio, A. (1999). *The feeling of what happens: body and emotion in the making of consciousness*. Harcourt Press, New York.

Damasio, A. (2004). Emotions and feelings: a neurological perspective. In Manstead, A., Fridja, N., and Fischer, A., editors, *Feelings and emotions*, chapter 4, pages 49–57. Cambridge University Press.

Darwin, C., editor (1872). *The expression of the emotions in man and animals*. D. Appleton and Company.

Ekman, P., editor (1982). *Emotion in the human face*. Cambridge University Press.

Filman, R. E., Elrad, T., Clarke, S., and Aksit, M., editors (2004). *Aspect Oriented Software Development*. Addison Wesley.

Frijda, N. (1987). *The emotions*. Cambridge University Press.

Gentner, D. and Stevens, A. L., editors (1983). *Mental models*. Lawrence Erlbaum Associates, Hillsdale, NJ.

James, W. (1884). What is an emotion'. *Mind*, 9:188–205.

Ortony, A., Clore, G. L., and Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge University Press.

Plutchik, R. and Kellerman, H., editors (1980). *Emotion: Theory, research and experience. Volume I: Theories of emotion*. Academic Press, New York.

Sanz, R., López, I., Bermejo-Alonso, J., Chinchilla, R., and Conde, R. (2005). Self-X: The control within. In *Proceedings of IFAC World Congress 2005*.

Sanz, R., López, I., Rodríguez, M., and Hernández, C. (2007). Principles for consciousness in integrated cognitive control. *Neural Networks*, 20(9):938–946.

Sanz, R., Matía, F., and Puente, E. A. (1999). The ICa approach to intelligent autonomous systems. In Tzafestas, S., editor, *Advances in Autonomous Intelligent Systems*, Microprocessor-Based and Intelligent Systems Engineering, chapter 4, pages 71–92. Kluwer Academic Publishers, Dordretch, NL.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5:42.

Trapl, R., Petta, P., and Payr, S., editors (2002). *Emotion in Humans and Artifacts*. MIT Press.