THE AUTONOMOUS SYSTEMS LABORATORY
UNIVERSIDAD POLITÉCNICA DE MADRID

A Model of Emotion as Patterned Metacontrol

To appear in Biologically Inspired Cognitive Architectures

Ricardo Sanz, Guadalupe Sánchez-Escribano and Carlos Herrera

ASLab A-2013-003 v 0.0 Draft

January 2013

Abstract: Adaptive agents use feedback as a key strategy to cope with uncertainty and change in their environments. The information fed back from the sensorimotor loop into the control subsystem can be used to change four different elements of the controller: parameters associated to the control model, the control model itself, the functional organization of the agent and the functional realization of the agent. There are many change alternatives and hence the complexity of the agent's space of potential configurations is daunting. The only viable alternative for space- and time-constrained agents —in practical, economical, evolutionary terms— is to achieve a reduction of the dimensionality of this configuration space. Emotions play a critical role in this reduction. The reduction is achieved by functionalization, interface minimization and by patterning, i.e. by selection among a predefined set of organizational configurations. This analysis lets us state how autonomy emerges from the integration of cognitive, emotional and autonomic systems in strict functional terms: autonomy is achieved by the closure of functional dependency. Emotion-based morphofunctional systems are able to exhibit complex adaptation patterns at a reduced cognitive cost. In this article we show a general model of how emotion supports functional adaptation and how the emotional biological systems operate following this theoretical model. We will also show how this model is also of applicability to the construction of a wide spectrum of artificial systems[1].

[1]This work has been partially funded by the European Commission through grants FP7-ICT-231453 HUMANOBS (Humanoids that Learn Socio-communicative Skills by Imitation) and FP7-PEOPLE-COFUND-2008 (Autonomous Cognitive Systems).

Keywords: Cognitive architecture, Functional organisation, Emotion, Adaptive Systems, Autonomous Systems.

An Approach to Emotion Research

Emotion theory is a very complex research domain due to the multifarious nature of emotions and the multiple purposes, views and implications of such a theory (Houwer and Hermans, 2010). As is the case with any scientific endeavor, cognitive science of emotion shall advance by means of hypothesizing theories against a backdrop of empirical data. In the context of emotion, however, this backdrop of empirical data is, at least, confusing. From skin conductance measures, FMRi scans of the amygdala, or concentrations of neuromodulators, to overt social behaviors, facial expression or verbal reports of emotional qualia, the variety of facts to take into account is daunting. It is difficult to categorize and integrate the *ground truth* that we want to explain in biological emotion.

In a parallel line of activity, autonomous systems engineering is looking for functional control architectures that could augment the performance and resilience of technical systems operating in open-ended environments. In this work thread, engineers are looking at the biology of emotional mechanisms in search for architectural inspiration to achieve the levels of robust autonomy that animals demonstrate (Fellous and Arbib, 2005; Samsonovich, 2012). Robustness is needed both for functional and safety reasons, since deployment environments and required functionalities are increasingly varied (Sanz et al., 2007a).

In this dual-context situation —biological cognitive science and cognitive systems engineering— *models* do play a central role. Models constitute the cornerstones of both theories and implementations of systems. In particular, computer models of emotion (Dyer, 1987; Marsella et al., 2010) play the role of rigorously expressing a theoretical model for evaluating it.

In this article we will present a concrete, cybernetic, vision on what emotion is and propose a derived agent architectural model that can serve as a blueprint for a *general theory of emotion*. This theory can be used for both a) biological explanation of emotional phenomena and b) engineering of better autonomous systems (Arbib and Fellous, 2004). The *Patterned Metacontrol* model of emotion is based on the basic idea of appraisal-driven metacontrol of functional organization patterns of the cognitive-emotional agent.

The rest of the article will i) describe the state of the art in emotion research from a systems modeling perspective, ii) analyze the context of adaptive systems where this work sits, iii) introduce the issue of reconfiguration-based adaptation in high autonomy systems, iv) propose the model of emotion as pattern-based metacontrol which is the main content of this article, and v) conclude with an evaluation and future perspectives for this work.

The Cognitive Modeling Flow

Computer implementations of cognitive models are common research assets in cognitive science (Sun, 2008; McClelland, 2009). These models are created to better understand a concrete biological or psychological process by both a) the rigorous study that is necessary for building such a computer model and b) by a testbed evaluation of the model, constituted by a running program. Computer models are built because the complexity of the modeled system is so high that performing mathematical studies of analytical theoretical models is inviable. Computer models can be simulated to validate them.

Figure 1 shows an overall activity flow when developing cognitive models using computer programming. These computerized models can be built from scratch as conventional programs or can often be implemented using programmable cognitive modeling engines like ACT-R (Anderson et al., 2004; Stewart and West, 2007) or SOAR (Laird et al., 1987; Laird and Newell, 1993) .

In many cases, the implementation and proper evaluation of computer models of some specific phenomena requires their framing inside wider realizations of general cognitive architectures. Specific implementations of general theories of mind are used to provide and adequate and complete experimental context, including all the cognitive features that are necessary for the evaluation of the more specific models of concrete psychological phenomena. For example, the evaluation of a model of an emotion like fear may require a minimal agent capable of both perceiving some frightful entity or situation and acting in accordance with what has been perceived. In a very precise sense, as will be addressed in this article, emotions make sense only over other cognitive competences of the agents.

The computer-based model development process, whether using a conventional programming language or any of the domain specific languages of these cognitive architecture engines, starts with the selection of a concrete cognitive aspect to be addressed; e.g. visual target tracking, spoken word recognition in cocktail parties, fear of bears, *etc.* This behavioral aspect will be analyzed and characterized in terms of model needs. The modeling workflow continues with the proposal of an abstract model that will be later translated into a more concrete realization: a computer implementation. Abstract models are usually described using textual, narrative descriptions. In many cases informal diagrams do accompany the descriptive text.

These diagrams are usually composed of boxes and arrows (Shaw and Clements, 1997) and help communicate the inner structure of the model. This may improve the precise understanding of the model, but in many cases it may convey a false impression of rigor, due to the concretion of the model in a componential structure and a set of causal signal flows. Note that the abstract model, while being abstract, must still be precise enough as to rigorously drive the model construction process (Bosse et al., 2008; Sanz et al., 2010b).

Obviously not all cognitive models described in the literature comply with this

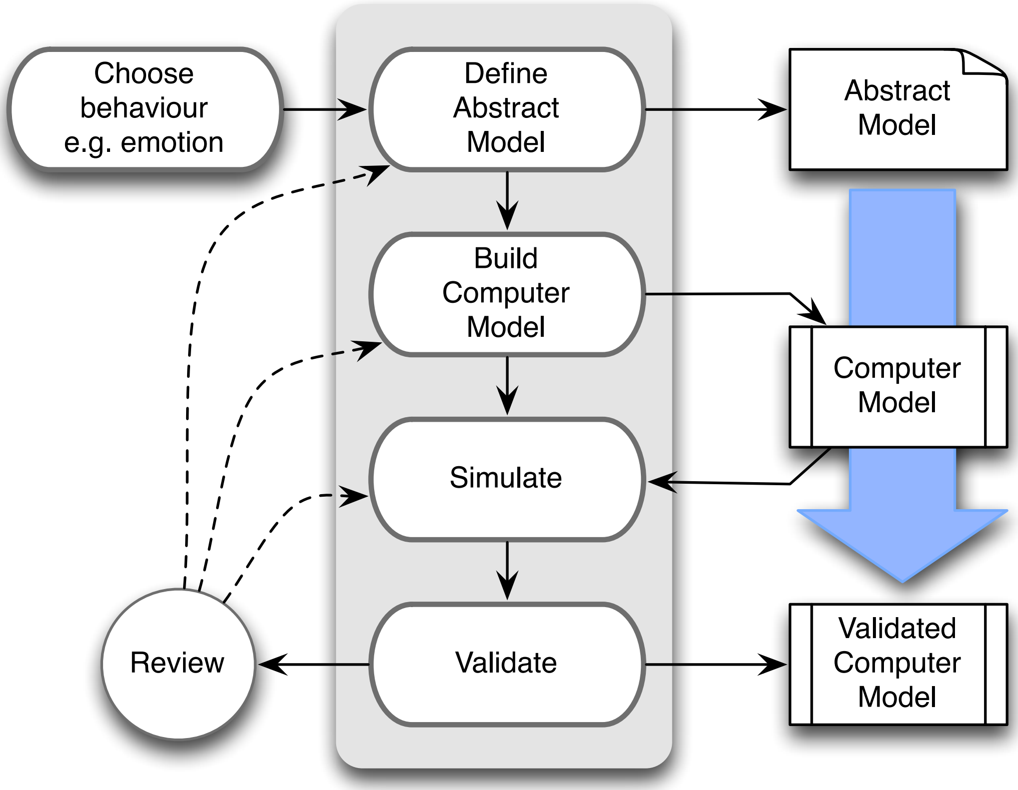

Figure 1: A common activity flow when developing cognitive models of concrete agent behaviors using computer programs: from the abstract model to the computer implementation. The central box shows the main modeling activities and the different models are shown on the right part. The abstract model is usually captured using non-formal languages. The computer model is formal and is validated by means of simulation.

precision requirement. Many of them are expressed in too weak terms to be easily translatable into programs. In other, even more extreme cases, the described models contain not just fuzzy but non-computable and/or metaphysical elements that are a-priori non-translatable to computer programs. Computer implementations are a filter of weak, non positive theories; a filter that may be considered too strict by some researchers, because it rules out the possibility of incorporating non-computational aspects into the models.

The step from abstract model to a concrete one requires the *building* of a computer model in the precise terms of the programming language used. This step is usually done by programmers codifying the abstract descriptions depicted in text, boxes and arrows by means of a computer programming language. This language may be a general purpose programming language like C, Java or Python, or a domain specific language as is the case of the programming languages of SOAR, ACT-R, NeuroML (Gleeson et al., 2010), or BRAHMS (Mitchinson et al., 2008).

Computer programs can only be executed if they are complete. However, the abstract models captured in text-and-boxes descriptions are not complete —in systems engineering terms. The computer model has to be completed —filled-in the blanks— with some magic from the hands of the programmer. Additions and hacks must be introduced in the program to address necessary issues not specified in the model —*e.g.* I/O relations, system maintenance issues, data formats and organization, *etc.* The computer model suffering this enlargement departs from the abstract one, sometimes substantially, as those hacks are usually not conceptualized and fed-back into the abstract model.

A more rigorous and complete process for cognitive model and architecture specification is necessary in order to fully evaluate and compare alternative models (Sanz et al., 2009). However, while this problem is pervasive in computer modeling of cognition —esp. in immature areas like emotion or consciousness— the modeling strategy shall be kept apace, as it is the core strategy of successful science.

The computer model built in this way must be *validated* and this is done by means of *simulation*. The specific simulations to be performed will depend on the concrete implementation of the model. In many cases the testing of emotion models is done in whole agents —simulated or robotic— to evaluate the behavioral responses in certain environmental settings (See Figure 2).

Experiments with virtual rats in virtual mazes are common trade in this context. The biological significance of these results is dubious, as the ecological relevance of the virtual contexts is under discussion. The experiments, however, are able to demonstrate that certain cognitive architectures are able to produce certain classes of behaviors that are homologous to biological behaviors at a certain level of abstraction (Webb and Consi, 2001; Webb, 2009).

Simulation results are validated against the ground truth. In this process there are two levels of validation that sometimes get mixed up:

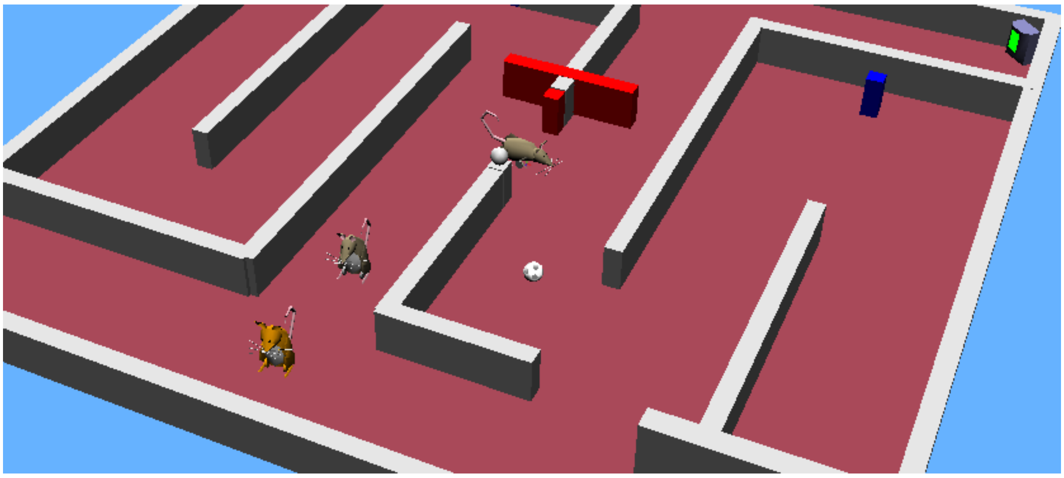

Figure 2: The VRat Webots-based simulation ground for the evaluation of the cognitive-emotional models developed in the ICEA research project (Hernandez et al., 2010).

- *Implementation validation*, *i.e.* checking that the programmed implementation complies with the abstract model.
- *Model validation*, *i.e.* checking that the abstract model fully addresses the emotional explananda.

If the simulation results are good enough, the validation analysis can be feed into the model to release a final, validated version. On the contrary, if validation results are not good enough they may lead to a necessary review of the model at any of the previous stages. This implies returning to previous stages of the modeling workflow:

- Perform new simulation experiments to get more data concerning other experimental environments, situations or statistically relevant data concerning agent behaviors.
- Reprogramming of the model when the behavior produced is not the expected one and the source of the problem is erroneous mapping to code.
- Revision of the abstract model if the previous revisions do not produce the expected results. This is the hardest and seldom done backtrack movement because it implies both abandoning previous theoretical thoughts and producing a new theoretical model —always a hard task.

In this process there is an enormous risk that is not fully recognized. When reprogramming the model, using the unconstrained hacking strategy mentioned before, it is always possible to obtain any desired behavior thanks to the flexibility of software[2]. The model and the program can get broken apart in this process, losing the necessary modeling connection. To avoid this rupture, the programed model and the abstract model must be kept in consonance. It is necessary to be aware of this problem and use development methods that can keep and enforce this correlation. The strategy used in the developments described in this article is called model-driven development (Balmelli et al., 2006; Sanz et al., 2009), where the step from abstract model to code is automated by means of development tools, hence leaving no gaps for model-program mismatch (Sztipanovits and Karsai, 1997).

Motivation for emotion in technical systems

As was said before, most computer models of emotion are built to evaluate theoretical models that cannot be verified but by simulation in virtual testbeds. However, the pursue of technological assets for the implementation of emotional systems in artifacts may have other, quite different motivations. For example Kushiro et al. (2013) show how emotion can be used to drive conceptual learning processes for robots in unknown environments.

The rationales for research on artificial emotion can be reduced to three basic kinds:

1. Evaluation of theoretical models of emotion. This is the main context, as was described before.

2. Implementation of human-like machines. This can be done for the improvement of usability of machines in human-tuned niches —*e.g.* to better handle human-machine interaction— or can be done for the simple hubris of building artificial humans.

3. Construction of better artifacts. The idea is to exploit architectural control patterns reverse-engineered from animals. Bio-inspiration in this context is guided by the evolutionary psychology dogma that all mental traits are there because of their adaptive value. In principle, they could be also leveraged in artifacts (Sanz, 2003).

The work described here is part of a larger research project whose objective is to produce universal control technology for improving mission-level resilience in technical systems in broad domains. The *ASys Project*[3] pursues the global objective of attaining *robust autonomy* by means of incorporation of cognitive mechanisms

[2]This risk is reduced when using constrained programming models as is the case of the domain specific programming languages and is maximal when programming using generic programming languages.

[3]http://www.aslab.org/public/projects/ASys

into the machines. Both terms —*autonomy* and *robustness*— do have precise meanings in this context (Sanz et al., 2000; Hansen and Sargent, 2007): being able to fulfill missions even under the presence of external and internal disturbances.

This article describes an initial theoretical result emerging from the on-going general design work that is leveraging architectural organizations of emotional mechanisms of animals in search for the needed robustness. Note that this work is not trying to build models to replicate biological behavior but to create universal architectural assets. This implies that the validation of the models is not going to be done in terms of biological ground truth —*i.e.* by animats imitating behavior of real animals— but in terms of absolute performance of technical systems.

The model presented, however, intends generality and, as such, may lead to implementations of ecologically-relevant experimental testbeds; hence serving the purposes of biological model validation as described in the previous section *The Cognitive Modeling Flow*.

Emotion in vivo and in silico

There are many facets of emotion that can constitute the ground truth that is the target of the modeling effort. Affective states, emotional states, cognitive impact, physiological impact, evolutionary development, emotional expression, emotional appraisal, *etc.* have been catalogued and used as raw backdrops for the wide variety of emotion theories.

In all this landscape of the emotional phenomena, we can identify four core aspects of emotion:

Behavioral/expressive — Generate specific modes of action and doing an externalization of inner state for social interaction.

Experiential/subjective — State valuation and generation of qualia associated to the appraisal of situations.

Somatic/neurophysiological — Changes in bodily state to properly react to situations; includes alterations in neurophysiological and hormonal systems.

Cognitive/interpretive — State representation of ongoing situations and impact to the agent; updating of the mental model about goals and values.

Considering the objectives of our research project, a fundamental question emerges: Are all those many aspects to be considered in machines? Or only the last one has any meaning? The answer lies on the teleonomic nature of both. Machines are very different from humans in many senses, but there are two aspects of maximal relevance for addressing these questions:

Purpose: Humans are natural and machines are artificial; this shall be understood in the precise sense of *artificial* defined by Simon (Simon, 1996). Humans are not created with a purpose —usually— but machines always are. The professional competences of the engineers are specifically focused on that: build systems that satisfy allocentric purposes.

Realization: The design of humans is evolutionary and their construction is organic. The design of machines can be constructionist or constructivist —somehow mimicking biological "design" (Chaput, 2004; Thórisson, 2009). The construction of machines is never organic but in limited experimental settings (Lipson and Pollack, 2000) or in the case of virtual machines (Serugendo et al., 2011; Tchao et al., 2011).

This implies that when trying to elaborate a general theory of emotion, purpose and realization cannot be part of the theory. A general theory of emotion shall be of applicability no matter what is the purpose of the system and independently of the realization technology. This may then imply that only the experiential/subjective and the cognitive/interpretive aspects are of relevance for an universal, architecture-centric theory of emotion.

Emotion in biological systems

Affective neuroscience has tried to identify the neural bases of emotion (Panksepp, 1998; Davidson and Sutton, 1995) following a research line opened by James Papez centered on the brain-wide distributed aspect of emotion. Papez (Papez, 1995) proposed to distinguish two fundamental aspects of emotion: the stream of thought and the stream of feeling. In the first, the stream of thought, sensory information is transmitted via the thalamus to the neocortex side areas to become perception, thought and memory. The stream of feeling is the basis of the feeling of subjective experience via thalamus to hypothalamus (in particular hypothalamic mammillary bodies), arriving in the cingulate cortex, which hosts the emotional experience. These pathways are represented in the popular "Papez circuit", a classical neuroscience boxology example (see Figure 3).

The Limbic System

Paul MacLean (MacLean, 1949) proposed a global emotional theory basing his work on theories of Papez, Canon, Klüver and Bucy. Heinrich Klüver and Paul Bucy (Kluwer and Bucy, 1939) conducted studies of the brain areas involved in visual hallucinations caused by drugs. They obtained important results on visual perception, long-term memory and emotion that strongly influenced the work of Paul MacLean.

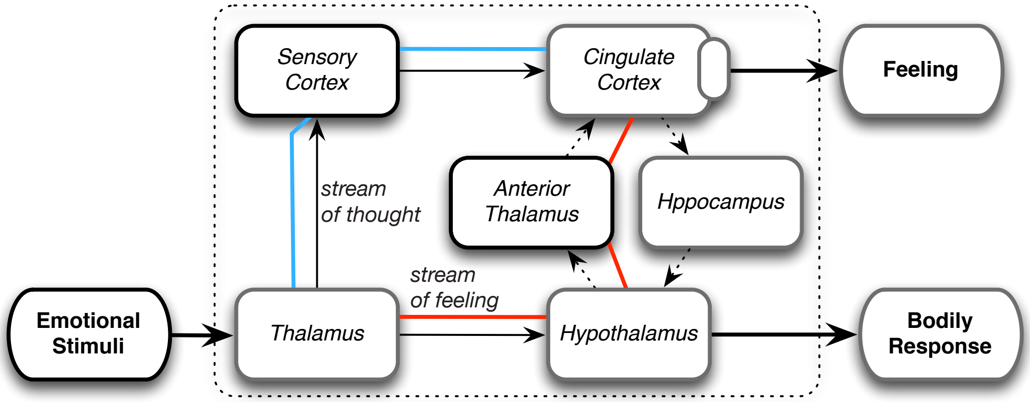

Figure 3: The Papez dual circuit theory of the functional topology of emotion in the brain (Papez, 1995). Both the stream of thought and the stream of feeling go from the emotional stimuli to the feeling. The "gyrus cinguli" —the cingulate cortex— acts as integrator of the thought and feeling signal streams to generate the emotional feelings. Bodily responses are triggered only by the stream of feeling.

According to MacLean, the brain contains the signal integrators underlying emotions that fuse perceptions from outside with gut feelings. In his triune brain theory, the proposed limbic system —septum, amygdalae, hypothalamus, hippocampal complex, and cingulate cortex— is the product of early phylogenetic evolution for survival and protection of the individual, hence being a major functional system for the agent.

Subsequently, various studies including LeDoux (1996, 2000a) have shown that the limbic system is not a complete structure well delimited functionally or anatomically. This has led, therefore, to eliminate the idea of unique neural circuits for emotion.

The Amygdala

LeDoux and also Phelps (2006) studied the amygdala and its role in fear conditioning processes and defense responses in rats. In the amygdala several signals converge both from external perception (auditory, visual, *etc.*) and somatosensory perceptions —*e.g.* of electric shocks or other unconditioned stimulus. Under this viewpoint, two sensory pathways converge in the same center in the brain —the amygdala: The conditioned stimulus crosses the thalamus and the cerebral cortex to the amygdala, and the reactions of the unconditioned stimulus through the thalamus and the somatosensory cortex. From the core of the amygdala information flows to the brainstem and hypothalamus to result in the behavioral and physiological responses of fear.

Several neuroimaging studies in humans (Adolphs et al., 1995; Phelps and LeDoux, 2005; Dolcos et al., 2011; Rosen and Donley, 2006) have established that the amygdala is activated in in proportion to the perception of fear. Morris et al. (1998) studied this same activation when the conditioned stimulus is unconsciously perceived, obtaining a similar result. The activation of other subcortical nuclei (superior colliculus and pulvinar) influenced the theoretical conclusion that there are subcortical ways to detect emotional phenomena. The amygdala plays an important role in emotional memory, in the recognition of facial and body expressions of fear, *etc.* See later section on *Patterns in Biological Emotion*.

The Prefontal Cortex

Harlow, Phineas Gage's physician, discovered an astonishing personality change following an injury in the prefrontal cortex. Subsequent work — *e.g.* from Damasio (Damasio, 1994) confirms the participation of prefrontal cortex in the brain systems involved in emotion. This is also confirmed by current human neuroimaging studies (Davidson et al., 2003). Apparently, it has only a modulatory function of additional components that are deeply involved in the generation of the emotional response. Miller et al. (Miller and Cohen, 2001) and Ochsner et al. (Ochsner et al., 2002) suggest that the prefrontal cortex helps adjusting the emotional response to be adaptive.

Kringelbach (Kringelbach and Berridge, 2009) states that the orbitofrontal cortex (OFC) of nonhuman primates is activated by primary reinforcers (*e.g.* food when the animal is hungry) and is deactivated in satiety situations. This suggests that the OFC is involved in learning emotional and motivational value of stimuli, as it is involved in the emotional response when it involves reinforcement-learned contingencies.

Emotions influence reasoning and decision-making behavior. Damasio proposes the theory of somatic markers to explain the relationship between emotion and reasoning. The Phineas Gage (Barker, 1995; Baars and Gage, 2007) case analysis has led to the conclusion that his personality change came from the lesions in the OFC and the VMC (ventromedial cortex). According to the theory of somatic markers, the processing of somatic signals of emotional nature has its substrate in the VMC and the medial sector of the OFC.

The VMC is a secondary association area that integrates information from different sensory modalities, somatosensory areas responsible for keeping the body bioregulatory equilibrium, areas involved in emotion processing, and also areas related to working memory. VMC and OFC play a critical role in the evaluation of external information about ongoing events, and internal information about bodily sensations and emotional state.

The Anterior Cingulate Cortex

Different studies (Bush et al., 2000; Davidson et al., 2002; Shackman et al., 2011) suggest that the anterior cingulate cortex (ACC) appears to be related to the expression of emotion and conscious experience (also associated by Papez with conscious experience). This work establishes a distinction between an affective section and a cognitive section of the ACC, occupying posterior and anterior parts of the area in question, respectively.

The affective section seems to be part of the autonomic nervous system activation control in emotional experience dealing with afferents related to emotional experience (nucleus accumbens, amygdala, *etc.*). The cognitive sector is closely related to cognitive processing, the selection of emotional responses and reasoning to solve conflict within the information available for decision making. It is closely related to the dorsolateral prefrontal cortex, posterior and parietal ACC and also the supplementary motor area and the spinal cord.

Searching Fundamental Emotion Theories

The ASys research program tries to leverage emotion in machines in a rigorous and architecture-centric, systematic way (Bass et al., 1996; Sanz et al., 2008). In order to do that it is necessary to have a theory of emotion that departs from biological constraints and is fully rooted in general systems theory.

The traditional approach to biologically inspired control architectures for machines is exemplified in the paradigmatic example of building robot controllers inspired by neuroscience results (Dario et al., 2005; Tamburrini and Datteri, 2005; Ziemke, 2008; Hernández et al., 2011; Krichmar, 2012). This is done by both neuroscientists and roboticists, each for the specific purposes of their own domains of activity. Both *robot-based cognitive modeling* and *brain-inspired robotics* try to leverage results and possibilities of the other domain (see Figure 4).

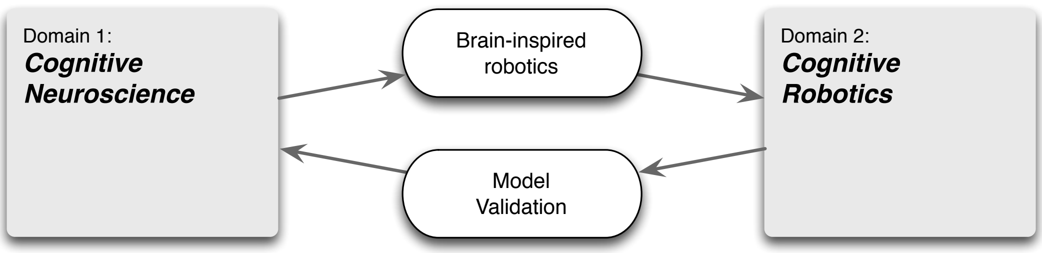

Figure 4: The conventional dual bio-inspiration paths from biology to technology and back. Cognitive robots are built following cognitive neuroscience models and are used as realizations to test those cognitive models. Biology inspires technology and technology evaluates biology.

The two activities exploit the potential cybernetic equivalence —*i.e.* the equivalence at a specific systemic level— between the bodies, sensors and actuators of robots and animals. This equivalence implies that a controller —a cybernetic engine for action— may be translated from one system to the other if some equivalence conditions apply. The most common strategy has been to create robots based on reverse-engineered functional and anatomic relations in the brains of different classes of animals (Hernandez et al., 2010). If the brain reverse-engineered, neural controller for the robot works, this means that the cognitive neuroscience theory was right and that it may be used to build robots.

As a paradigmatic example, researchers have looked at man and tried to create a robot in his image and likeness, in the hope that a human-like embodiment will render the neural architecture meaningful. This is the core argument and strategy of humanoid robotics and biorobotics (Webb and Consi, 2001; Aldridge et al., 2000; Breazeal, 2004; Gini et al., 2009).

However, this strategy is strongly misguided, and most of the results shall be considered as only relevant for the impact they may produce in mass media (Sanz et al., 2010a). The implemented architectures are usually tuned to the concrete experimental conditions of the tests (*e.g.* rats in virtual labyrinths, see Figure 2) and lack the flexibility of real neural controllers and the dependability needed in robot applications.

There is no easy engineering method to apply these architectures to real systems; generality and robustness are missing. The only possibility left by this approach is the systematic evaluation and testing in *all operational conditions* to get a statistical guarantee that the robot will behave properly. This verification and validation strategy by means of testing is similar to the strategy used in some software applications (*e.g.* productivity suites or video games) but it is useless for systems of higher operational complexity or tighter safety requirements. A more adequate research strategy is depicted in Figure 5.

The transfer of control designs from biological realization in the brain to real-world, dependable robots must pass through a theorization process. This means that we shall 1) *extract generic control architecture designs from the brain* to be able to 2) *apply them rigorously in a robot engineering process*.

To this end, and specifically in relation to leveraging the value of biological emotion, the ASys research project follows an stepped program:

1. Find fundamental emotion architectural organizations in the brain.
2. Eliminate unnecessary biological details in the modeling of emotion.
3. Consolidate these organizations in domain-independent models.
4. Transfer to realizations of technological usability.

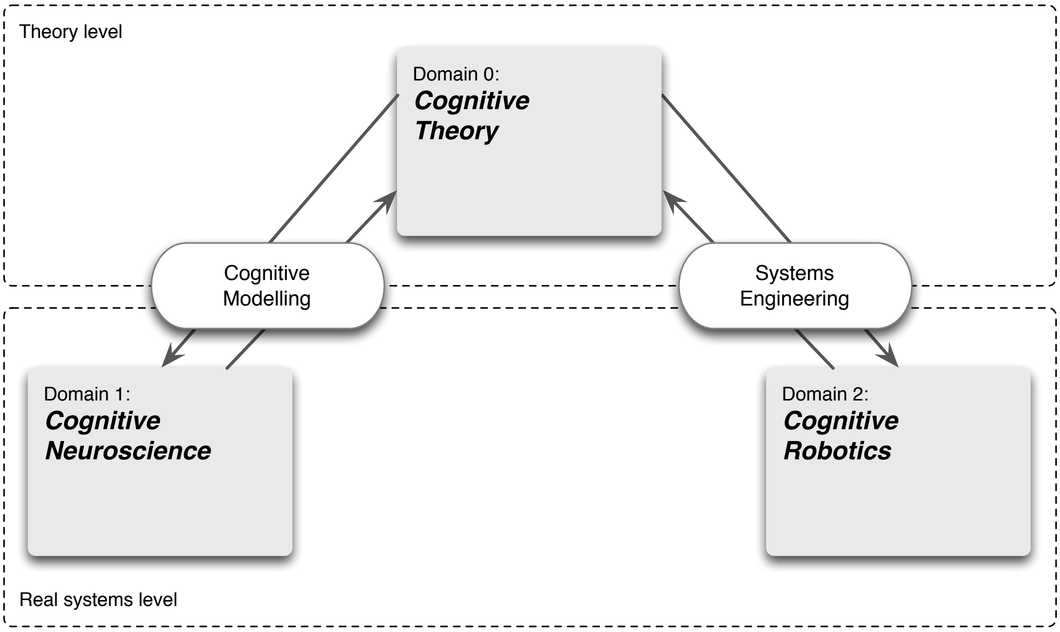

Figure 5: The path from cognitive neuroscience to cognitive robotics shall go through a general theory of cognition. Cognitive robots shall be built using rigorous systems engineering methods and not as cognitive models. Cognitive modeling shall produce rigorous theory and avoid falling into the easy methods of computer and robot melling.

As was said before, the variety of the emotional phenomena makes difficult the elaboration of grand, unified theories (Moffat et al., 1995). The emotion phenomena seems too heterogeneous to be addressable by a single theory. However, emotion theorists have tried to do so. The usual strategy is to identify the *core aspect* of emotion and propose a theory for it. This theory can latter be applied to the derivation of ad-hoc explanations for specific emotions and emotional scenarios.

These theories may have many different forms, as is the case of *spreading activation models* —networks of nodes of cognitive-emotional bundles—, *componential models* —organizational variations in cognitive processes—, or *parameter-based models* —appraisal dimensions that map to expressions. These theories span and are scattered along the whole emotional process, from appraisal to external effects of emotion. The theories of Lazarus (Lazarus, 1991), Panksepp (Panksepp, 1982), Damasio (Damasio et al., 1996), LeDoux (LeDoux, 1996), Scherer (Scherer, 2009) and Rolls (Rolls, 2005) are good examples of deep theories trying to deal with emotions in this way.

Adaptive systems

The theoretical framework for the ASys theory of emotion is the framework of autonomous systems. Autonomous systems are able to reach their goals in the presence of disturbances. The more complex the goals and the environment are, the more sophisticated the agent should be. Control and cognition are the solutions that we explore. Complex cognition is required to build advanced autonomous systems. Ashby law of requisite variety (Ashby, 1956) states it clearly: the variety of the controller must match the variety of the problem.

Agents and Value

A characterization of the nature of goals and the way goals are attained by the agent seems necessary in order to understand our model of emotion. As we will see, emotions and agent goals are closely related through the concepts of *value* and *function*. In essence, emotions will shape agent's functionality in order to reach valuable goals. Cognitive-emotional agents can determine their own functional organization.

Agents having advanced cognitive capabilities —*e.g.* Dennett's Popperian agents (Dennett, 1996)— will not merely react to the present state of things but take also into account what the future will be. Agents anticipate the future state of their world (Rosen, 2012; Lorini, 2011). They will decide about their many potential actions taking into account how the future will be.

At any particular time instant, the selection and execution of a concrete action alternative will shape the agent's future. To properly make a selection among the several possible actions, the agent shall not only anticipate its instantaneous effects or the final state that may be reached. It must also determine the evolution of a state-dependent *value-for-the-agent* along the whole, anticipated, system trajectory. The emotional system will help determine the value of each state of affairs, both present and anticipated. Note that due to the multiplicity of possible actions and to the pervasive uncertainty, there will be many potential, anticipated trajectories. The evaluation is hence not simple and straightforward, but implying several counterfactual, anticipated states.

This evaluation of expected value *along* the different alternative potential future trajectories is what the agent uses to decide what to do. This means that the final state to be reached is not strictly the *real goal* that autonomous agents seek. What agents seek is to have a behavior that maximizes the accumulated value got while behaving. As we will se later, emotions will play a critical role in shaping this behavior by shaping agent's organization (Marinier III et al., 2009).

In the case of biological systems, this machinery for future valuation may be embedded into the machinery for action generation, hence the agent lacking any explicit form of valuation representation. In conscious decision-making, however, this explicit representation is needed. This is the role played by the phenomenology of emotion (Sanz et al., 2012). We need explicit representations of the instantaneous state-of-affairs value for the agent to include them in cognitive control. This does not necessarily mean that any cognitive anticipatory agent must experience emotional qualia as we do, but the existence of such a phenomenology may be a need.

In summary, action selection generates behaviors that try to maximize accumulated value, not just reach a specific valuable state. The final states are hence conditioned by the shape of value along the trajectories across the agent state space. Emotions play a critical role in determining both the value of a specific state and the behavior of the agent guided by this value. We cannot understand how agents pursue goals without the elicitation of this emotional mechanism.

Emotion and Autonomy

In the pursue of a general technology for autonomy we are investigating the emotion-cognition integration architecture in biological agents. The rationale for this research is that emotions are adaptive —in the evolutionary biology sense— improving competence of the agent to operate in a certain, unstable niche. A dogma of evolutionary psychology is that all mental traits have or have had adaptive value (Cosmides and Tooby, 2003).

The systems-level analysis of the emotional phenomena led us to two architectural conclusions:

- Emotions are control architecture systems directly serving the goals of the autonomous agent.

- They do so by providing real-time adaptation mechanisms to the agent.

The two key words in this analysis are *goals* and *adaptation*. Emotions are mechanisms that augment the capability of the agent to focus its action readiness on goals of relevance. In every time instant, a complex agent may have many concerns: such as be warm, seek food, escape predator, *etc.* In essence, we can say that emotional systems are used to determine what, among the many things that the agent is concerned about, is important. Emotions do rank goals.

But beyond goal ranking —the well-identified appraisal processes of emotion— emotional systems do positively act towards those goals. And they do so by means of adaptation of the functional structure of the agent.

The issue of adaptation is pervasive in autonomous systems. Evolutionary theory focuses on adaptation processes at a scale beyond the individual (Gould, 2002). Morphofunctional theory focus on adaptation at the individual level (Hara and Pfeifer, 2003). Both natural and artificial autonomous systems do adapt at these two scales: animals as species and as individuals and machines as artifacts and product lines.

From this perspective, adaptation shall be framed in the context of autonomy (Sanz et al., 2000). To understand, explain and engineer autonomy we must consider three interrelated aspects:

$$Autonomy = F(body, environment, task)$$

We cannot say that "a system is autonomous" but instead "a system is autonomous to carry out a particular task, in a concrete environment with a specific body realization". If we change any of these factors we may lose autonomy to perform and attain the task goals. Time —*e.g.* a deadline— is also a factor to be considered but, in general, this is not an issue in the vernacular use of the term "autonomous" and in a more technical use of the term it shall be included in the specification of the task.

Emotion theory shall be framed in this context and we can say, summarily, that emotional systems appraise the state of the external and internal environment to determine its value in relation to a set of goals and reconfigure the agent organization to maximize the expectancy of value. If the analysis is so compact and simple, the questions are then: How is this realized? Why so many emotions?

Reconfiguration in Cognitive Systems

Reconfiguring Systems

System adaptation by reconfiguration is a very old topic in all classes of computing systems research; from fault-tolerance to modern embedded and web-based systems (Deconinck et al., 1994; Weston, 1999; Fiadeiro and Lopes, 2010; Bayar and Yurdakul, 2012). The backpropagation algorithm (Rumelhart et al., 1986) offers a strategy for reconfiguration in architectures using the neural network paradigm. Selfridge's Pandemonium (Selfridge, 1958) offers a strategy for tuning the pandemonium based on signals flowing back from the cognitive demons layer into the computational demons layer. The functional reconfiguration approach of de la Mata (de la Mata and Rodríguez, 2010) offers an universal model-based approach for process reorganization.

In general, reconfiguration is driven by processes that try to match the organization of the agent to what is needed to accomplish certain goal. Andrew uses the term "significance feedback" to refer to these processes (Andrew, 2009). We have used the term "structural feedback" for similar purposes (Sanz and López, 2000). In essence the issue is the change of the *function* of the system as realized internally to match the *function* of the system as required externally. Note the two different uses of the word "function" in the previous sentence (Lind, 1990).

The strategy that attains a maximum impact in the shaping of behavior with a minimal effort, is the change of the control systems of the agent. Sliding mode control (Edwards and Spurgeon, 1998) —SMC— or variable structure control (Zinober, 1994) —VSC— are forms of nonlinear control developed as evolution of linear controllers to address their operational range limitations. In these methods, the core nonlinear dynamics is produced by application of a switching control signal that forces a behavior that crosses a set of more "normal" dynamics. The control law, usually a state-feedback controller, is not a continuous function of time but a patchy one —continuously moving from one smooth patch to another.

In essence, the abstract structure of the controller —i.e. its control law— depends on the region of the state space where the system is. The controller switches control laws when crossing the borders of these regions. A common implementation is based on the fast switching among linear controllers at possibly very fast speeds. However the elementary controllers need not be linear and only continuity is required.

These controllers apply a strong control action —with practically infinite gain to force the behavior of the dynamic system to trajectories sliding along the restricted mode subspace— and must be applied with more care than other forms of more smooth and moderate nonlinear control. Due to design and implementation errors, the aggressive variable structure control action can lead to chatter, noise, energy losses, excitation of unmodelled dynamics and faults.

Change and Timescale

In the previous analysis, however, it is not clear whether we are talking about learning, adaptation, reconfiguration, tuning, etc. when talking about non-static control structures[4]. Apart of what parts of the system are changing, the temporal aspect seems critical in all these phenomena. Here we hypothesize that the main difference between cognitive control, learning and emotion is the fact that the reconfiguration provided by emotion operates in real time (Liu, 2000).

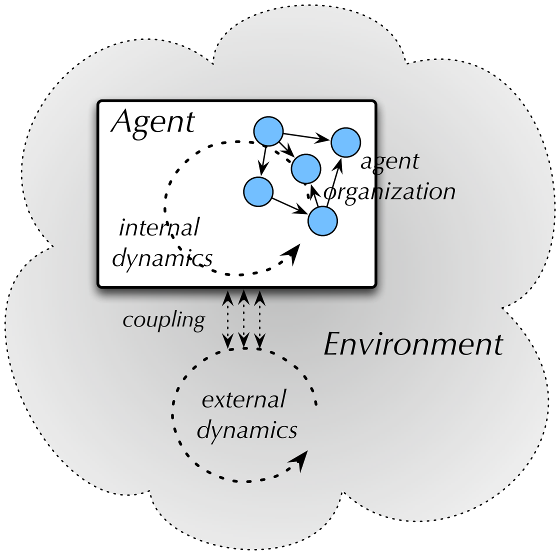

Figure 6: The dynamics of the agent-environment system is governed by the coupling of the dynamics of the world and the internal dynamics of the agent. Agent behavior is an emergent result of this coupled dynamics.

Agent behavior if not fully determined by the agent architecture but by the whole agent-environment system. The dynamics of the agent-environment system is governed by the coupling of the dynamics of the world and the internal dynamics of the agent (see Figure 6). The internal dynamics of the agent is determined by its organization —a bundle of parts and their relations(Rosen, 1970). Agents control their behavior by modifying their internal state and organization. This happens dynamically in search of value-for-the-agent maximization.

There are many processes at many different timescales in the dynamic change of agents:

[4]The simple flow of perception is also a dynamical change. However, we may assume the persistence of some structure in this process in the very sense of Klir hypothetical structure (Klir, 1969).

- Agents *react* to events in the environment.
- Agents cognitively *control* inner and outer processes to attain goals.
- Agents *learn* how to do things.
- Agents *adapt* to changing environments.

We can identify two levels of change: changes in the agent state and changes in the agent organization[5]. Conventional control systems engineering focuses in systems that change state but not organization. The key for general autonomy is *structural adaptation*: the change in systems organization to respond to environmental challenges. Some control systems technology is already addressing this but in a very limited way (*e.g.* some forms of adaptive control, context awareness or fault-tolerant control).

Structural Adaptation

By structural adaptation the agent changes its systemic structure. A system is defined by characterizing its parts and the relations between them:

$$Agent = \{Parts, Relations\}$$

In the case of cognitive agents like robots or animals we may identify two big classes of parts: physical parts and control parts. The changes that the agent suffers —reaction, control, learning, adaptation— may affect both the body and the mind. Mind structural changes are easier and more flexible due to the plastic, informational nature of minds.

When an animal decides between fighting or flying, there are no big physical changes in it. The fight-or-flight response —as hypothesized by Cannon— is a reaction to immediate threats with a general discharge of the sympathetic nervous system. The way of doing this is by *priming* the animal for fighting or fleeing. This *priming* is essentially a *priming of a specific functional organization of the biological system* of the animal. The sympathetic nervous system alters the *relations* between the animal *parts*. The concrete instantiation of the functional organization is selected dynamically by the activation of the sympathetic system and the adrenal-cortical system by the hypothalamus. These have cascading activation effects over the glands, muscles, adrenal medulla or ACTH (adrenocorticotropic hormone) system. The global behavioral disposition of the animal changes. The concrete response that is enacted —fight-or-flight— may then depend on minimal appraisals.

[5]Obviously a change of organization is a change of the agent state in a general systems sense, but we will use the term *agent state* to refer only to an internal representation of some state of affairs relevant to the goals of the agent (it is indeed a subset of the general agent state).

This selection of a functional organization is sometimes called *action readiness*. Action readiness is a core topic in behavioral emotion theories; for example Frijda (Frijda, 1986, 2007) specifically proposes that the behavior in emotion comes from animal's action readiness.

In a quick analysis, we consider that the most effective strategy for an economic agent is to perform the structural adaptation by means of changes in the structure of relationships and not in the parts that compose the agent:

$$Agent = \{Parts, Rels\} \rightarrow Agent' = \{Parts, Rels'\}$$

This implies that in general, what we can expect is that the agent addresses adaptation by the implementation of control mechanisms over the inner relational couplings.

Architectural Change

This model of *functional reorganization* has the required generality (see previous section *Searching Fundamental Emotion Theories*) for the purposes of our research. By focusing on systems architecture it is possible to address the analysis of multi-goal, teleonomic aspects of natural and artificial systems by means of their self-adaptation (Cheng et al., 2006). We will use the term "architecture" to refer to the functional organization of the agent (Bass et al., 2013; van Lamsweerde, 2003). Note that in general this functional organization includes both physical and informational parts; each one playing its specific role into the architecture.

Architectural change in the process of modifying the functional organization of the agent. It offers the possibility of creating flexible and resilient artificial agents that are better tuned to changing environments. It is a capability shown by brains when adapting to physical changes —e.g. due to strokes or surgery. This capability is exploited in adaptive robotics at different scales: from the micro-grained changes of neural evolutionary robotics to the macro-grained changes of fault tolerant industrial robotics. It is the plasticity of the architecture what is most critical in the provision of adaptive performance.

By architectural change we can create new agent organizations that are radically better tuned to their environments and goals. This means that architectural change is a *radical force for autonomy*.

Sometimes the analysis and/or design of a functional structure needs not addressing the whole system architecture. Some functionalities can be ascribed to partial organizations that can be *inherited* (or *reused* in the artificial world) by descendants; hence becoming organizational memes for species of agents. We will use the term "architectural patterns" to refer to them, borrowing the use of the expression from the domain of design of software systems (Gamma et al., 1995; Alexander et al., 1977; Cloutier and Verma, 2007).

Design patterns are functional organizations and reusable strategies used in software engineering. Architectural design patterns are design patterns that address high level organizational aspects of the system and are used to systematically design systems that manifest certain properties. System architecture is what determines most the performance of a system.

Patterns can be composed to get a composite of the features that they individually provide. We will look at emotion as an architectural pattern operating over other architectural patterns focused on action generation. We see emotion as a pattern-based pattern. Agents use emotions as mechanisms to select from a catalog of agent organizations and enact the most adequate to respond to ongoing challenges.

Taming the adaptation problem

Architectural change is a radical force for adaptation. Radical forces usually have associated difficulties and risks that shall be addressed to obtain viable systems.

Consider the problem of adapting a cognitive agent organization to an environmental change. There are plenty of possibilities of change that lead to myriads of alternative organizations. The space of architectural variation may be enormously complex (Sloman, 1995; Brun et al., 2013). The big risk is that almost all the possibilities will be wrong alternatives.

There is a big difficulty of a general adaptation strategy based on structural change: the dimension of the space of organizational possibilities is too big. We need criteria to decide what organizations or changes are adequate and what are not. In essence, to change the organization is to re-design the system and design is, in general, an inverse problem of enormous difficulty. This is so because the connection between architectural organization and goal-oriented effectiveness is extremely complex in general.

There are two general strategies that are of applicability for real-time solution of this problem: evolution and pattern reuse. In evolutionary approaches the new organization is produced more or less randomly —by mutations— and then selected by filtering criteria. It is used to see if it works or not, and is accepted if successful and rejected if wrong (Nolfi and Floreano, 2000). In essence, since we cannot go backwards from ends (goals) to means (designs) due to complexity, we try to go forward randomly to see what happens. This strategy is good to evolve animals if we don't mind that in most cases they are going to die due to dreadful mutations. But this is unacceptable for a commercial plane that must confront an engine failure.

The second alternative for architectural change selection is pattern reuse. This approach is based on reusing ways of organization that worked in the past. Instead of selecting random changes we restrict the selection to a *small group of previously* *known alternatives*. It is possible to simplify the design problem of autonomous agents by means of the generation of operational patterns: forms of agent components and partial organizations that are tuned to specific niches and tasks. Many fault-tolerant systems operate this way. Alternative patterns are identified at design time and realized in latent form until needed to address run-time faults.

The evolutionary and the pattern-based architectural adaptation methods are not incompatible. In fact, what we are suggesting in this article is that selection operates on functional patterns and not on whole agent architectures; i.e. patterns and not agents is what evolution selects.

Two basic "design" processes

The evolution of architecture-driven, functional reorganization of agents that implements the emotional mechanism is driven by two concurrent processes of evolutionary design:

Functionalisation — that produces systems composed by functional units.

Patterning — that captures and reuses subsystem organizations.

Functionalization is a process of function-based system re-organization. It is a form of refactoring the functional structure of a system. It usually operates under an evolutionary temporal horizon, but in plastic systems (e.g. those concerning mental functions in natural systems or software-intensive artificial systems) it may involve a single individual. It is driven by 1) the *modularization* of subsystems (e.g. the organs) by means of a process of subsystems' interfaces minimization, and 2) the *aggregation* into higher structures by means of interface homogenization i.e. the use of a common integration infrastructure (e.g. signaling molecules or computer protocols) that is shared across many subsystems.

The patterning process captures architectural patterns in a form that enables its reuse. The form of reuse will depend on the ontogenetic and adaptation processes of the agent. These patterns will appear both at the realization —i.e. anatomical— level and at the organization —i.e. physiological— level.

The relation between anatomy and physiology is the old relation between structure and function (Gómez et al., 2008). The many physiological patterns that are required for the different operating modes of the agent are multiple-intanstiated over the anatomy of the agent. The several subsystems of the agent (e.g. organs) can play different roles in different patterns.

Emotion as pattern-based metacontrol

At the end of a functionalization + patterning process we will have an agent that has a potentially large collection of functional patterns. In a specific instant there may be many of them that can be enacted over the actual anatomical substrate of the agent. A question hence arises: what patterns to activate at a particular moment?. The decision and the enaction of the functional patterns is the mission of the emotional system.

Deciding how to behave

The decision concerning what control patterns shall be activated depends on an appraisal of the agent-environment situation in terms of agent goals. The result of the appraisal is used in the instantiation of a pattern or patterns that best suit the ongoing situation. In the words of Scherer (Scherer et al., 2001) this is a process of:

> ”interrelated, synchronized changes in the states of all or most of the five[6] organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism.”

Scherer sees emotions as episodes of subsystem synchronization driven by nonlinear appraisal processes (Scherer, 2000). This is a view of emotion strongly linked to the component process model of emotion (Scherer et al., 2010). As was said before, this view is grounded on the dynamically enacted functional processes of the agent, hence being an ideal model for identifying emotion in cognitive architectures. For example, Ritter identifies some "emotional" patterns in a cognitive architecture like SOAR (Ritter, 1993).

The systematic study of emotion using an architectural approach must clarify what are the patterns that are enacted. This is a very difficult task for non trivial agents and/or non trivial emotions. Organisms with sophisticated emotions seem to be high in the complexity ladder. The use of a structure-centric, analytical approach is difficult due to the homogeneity of the integration infrastructure of the agent and the overlay of functional patterns over the same anatomic components. It is necessary to consider a behavioral analysis to complement the structural one. The combination of the categorical and componential approaches to the analysis of emotions may provide an adequate roadmap to pattern identification. As Panksepp says (Panksepp, 1998):

> ”The categorical approach can identify basic operating systems that exist in the brain, and the componential and constructivist approaches

[6]Information processing, Support, Executive, Action and Monitor.

> can provide schemata of how the genetically endowed systems develop their full resolution by interacting with the vast complexity of the real world."

In summary, we understand emotion as a real-time adaptation mechanism based on the activation of predefined patterns of organization of the agent. These patterns have been produced (evolutionarily) for specific activities/niches. In a sense, emotion selects the best available agent architecture for the present moment (esp. the control, informational part).

Emotion and Structural Feedback

The emotional system is, in essence, a control system. However, instead of being a controller in the domain of the physical magnitudes of the agent —e.g. temperature or molecular concentration— it is a controller in the domain of functional organization. Emotion implements a *structural feedback* mechanism (Sanz and López, 2000). In this control loop, the error signal is the evaluation of the degree of fitness between the functional state of the agent and the current state of affairs i.e. how well the agent is doing. The action affects the enacted system architecture.

In general, the degree of wellness is a too much faceted aspect as to be useful as a substrate for the patterning process. It is hence compacted before being broadcast as a synthetic appraisal state. This broadcast state is the emotion that is actually shaping the agent; an emotion that may be conscious and hence felt or may be not. This model of emotion implies that emotions can be subconscious.

The signals that compose the globally broadcast state —several appraisal dimensions— perform system-wide adaptation "to achieve a multi-level communication of simplified but high impact information" (Fellous, 2004) throughout the agent.

Emotions are system-wide signals driving the pattern-based reconfiguration process. In a sense, these emotions, as signals, constitute an emotional metacontrol bus (Sanchez-Escribano and Sanz, 2012). When represented at the level of awareness they constitute the emotional feelings. This vision of emotional experience is in line with higher-order thought (HOT) theories of consciousness (Rosenthal, 2009) and with the model of emotion proposed by Sellers (2013) that defines emotions as the mental experience of underlying motivations. Note that motivations are indeed instantiated by references to feedback control loops.

The emotional system is a control system controlling the organization of the other systems of the agent (Herrera-Perez and Sanz, 2013). In most cases these systems are themselves controlled and hence the emotional system acts as a metacontroller. The metacontrol bus is the set of signals that drive the reconfiguration, by perceiving the functional state of the agent and by acting over the functional elements of the agent. Figure 7 shows the deployment of an emotional system over

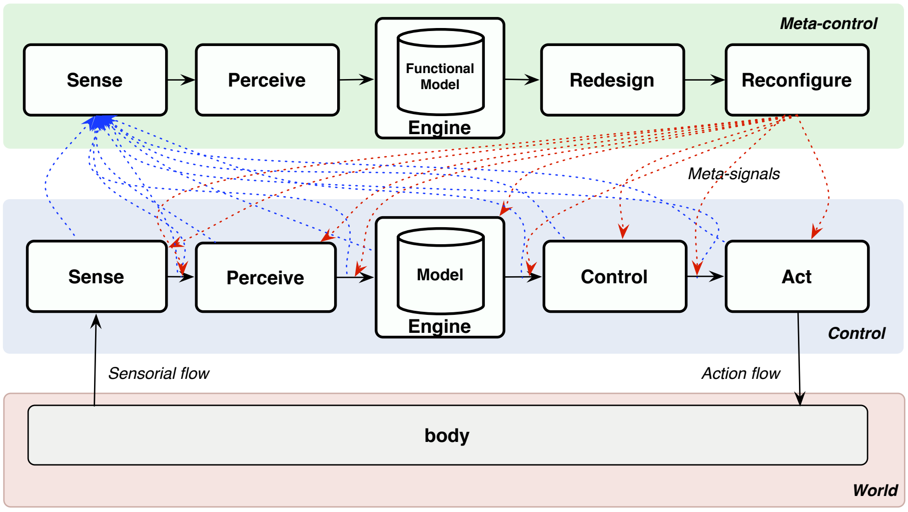

Figure 7: The emotional system can be seen as a control system controlling the lower control systems of the agent — a metacontroller. The metacontrol bus is the set of signals that drive the reconfiguration.

a cognitive loop based on the ECL pattern (Sanz et al., 2010c). This is a control system with two cognitive loops: the control level and the metacontrol level. The control implements a target domain closure: it controls domain magnitudes. The metasystem implements a morphofunctional closure: it perceives and controls system function. Figure 8 shows the activities coupling perception and action at the two levels. In essence, all the emotional system implements a metacognitive loop.

According to the conceptual act theory of emotion (Wilson-Mendenhall et al., 2011) it is the mental representation of the ongoing situation what determines the emotion experienced. In our theory we consider that the conceptualized situation is the functional coupling of the agent with its environment. Emotions are hence conceptually related to the functional structure of the agent. The set of concepts behind them provide an ontology that constitutes the semantic substrate of the morphofunctional elements of the agent (Herrera et al., 2012; Famaey et al., 2010). The concrete realization of these concepts and their dynamics can have multiple forms; see for example the dynamical systems realization by Treur (2013) or Larue et al. (2013) in this volume or the analysis of emotion-generating networks in the brain using a constructionist framework by Lindquist and Barrett (2012).

Emotion and Phenomenology

In essence, emotion captures the pursue of goals by leveraging morphofunctionality. At the top level of the agent goal system, emotional phenomenality provides the mechanism for assessing the relevance of the state of affairs. This is the reason

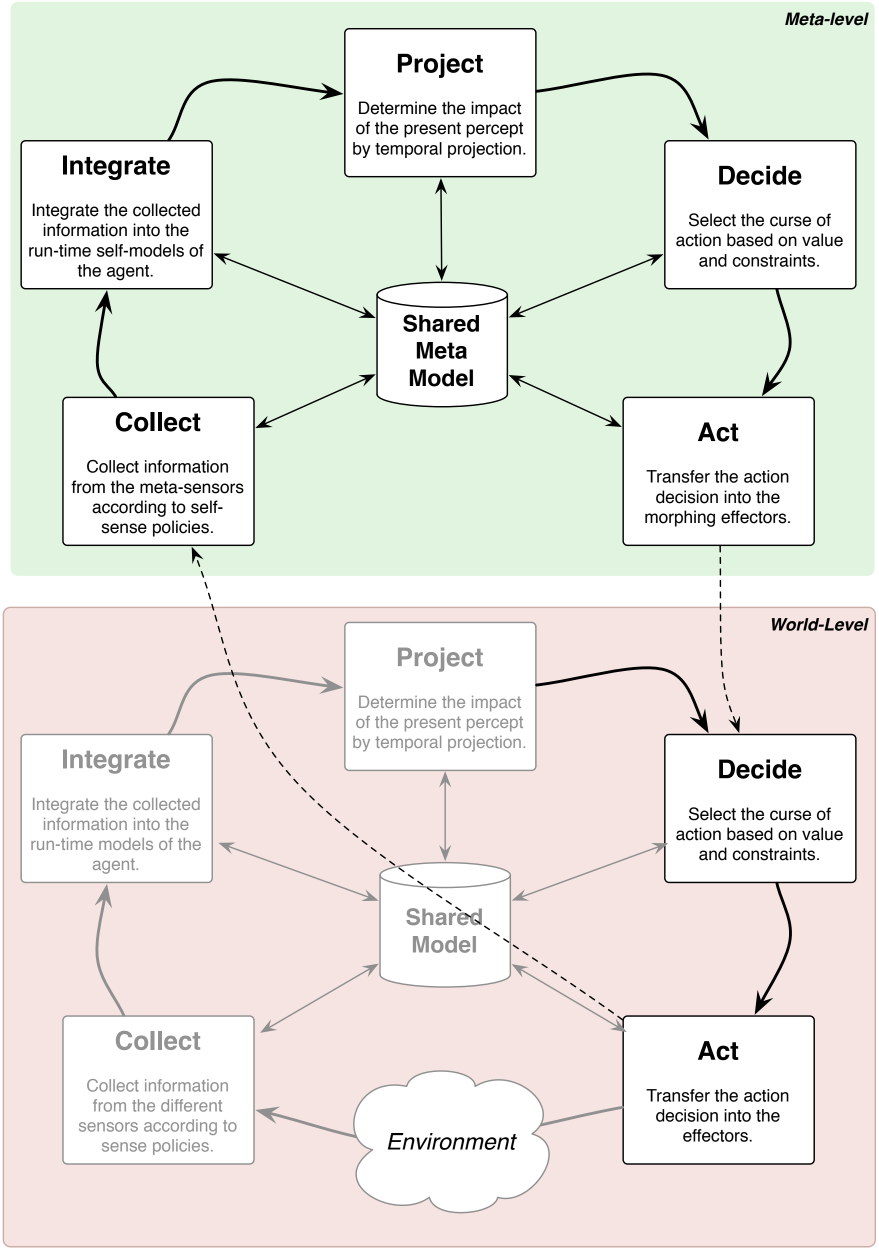

Figure 8: Cognitive metacontrol seen as the layering of two cognitive control loops: the world-level control implements a target domain closure, the meta-level control implements a morphofunctional closure. The lower level i) collects information about the target control domain (e.g. the robot environment), ii) integrates this information into the run-time model used by the agent, iii) projects into the future to anticipate states and determine value, iv) use the anticipated value to select the right action and v) executes the action. In the metacontrol loop what is observed is the functional state of the agent —sees what it is able to do— and acts by reconfiguration —changing what it will be able to do.

why emotions are so powerful: they force the agent to pursue the most important goals (Sanz et al., 2012).

When building emotion-experiencing machines, engineering the right phenomenological mechanism is crucial because it will be the origin of the intrinsic motivations of the agents. We shall adopt an heterophenomenological engineering approach to build phenomenologies into machines to match our own needs and not machine-specific ones (Bartneck, 2002).

Science and technology

Model evaluation is always a complex aspect of cognitive systems modeling. The evaluation shall be done in terms of what is the purpose of the modeling effort. In general, most emotion modeling theories are focused on the provision of models of the biological emotion. However, the model of emotions that we have presented here shall be evaluated in terms of both science and technology:

- Science — Its explanatory power of the phenomenon of natural emotion.
- Technology — Its potentiality as an architectural asset for building better machines.

For the first class of evaluation the theoretical model must address two dual aspects of emotion: 1) specific traits of animal or human emotion: appraisal, cognitive integration, emotion regulation, conscious and unconscious emotions, etc. and also 2) the concrete emotions themselves: fear, rage, jealousy, etc. Lots of painstaking work is needed to identify the metacontrol systems realizing these processes (Panksepp, 1998).

For the second evaluation task —artificial emotion— the development of the model must offer a path into technological consolidation in the form of both designs and reusable assets. These designs and assets, if adequate, could then be used in the construction of emotional systems for artifacts to improve their behavior. This improvement will essentially have the form of an increase in mission-level resilience and could then be measured by means of suitable benchmarks. Note that the emotional technology we are talking about is not a technology for emotion expression or for understanding of human emotional expression (Picard, 1997) but a technology for more robust machines.

Patterns in Biological Emotion

The patterned metacontrol theory of emotion can be used to frame and explain biological emotion. In this section we will use an specific emotion —fear– to investigate the role that the theory may play in the identification of control patterns in fear.

Fear is an adaptive response to stimuli that points to danger or threat, producing escape responses from these stimuli and thus from their potential hazard. Fear may also play a role at the multi-agent (social) level. Eibl-Eibesfeldt (Eibl-Eibesfeldt, 1971) suggests that besides being an adaptive response of single agents it also facilitates the emergence of social ties in collective defense.

There are four characteristics of fear to take into account:

- The triggering stimuli (any danger stimulus, undesirable variation of any stimulus, unexpected stimuli, etc.),
- the (maybe cognitive) evaluation and assessment of the situation and the state,
- the modulator factors influenced by the learning and socio-cultural context, and
- the functions: defense activation, emotion expression, decision-making, etc..

Usually, the triggering stimuli will come from outside the agent, but not always (just consider the fear triggered by an ongoing myocardial infarction). In any case this corresponds to the gathering of information at the lower levels. Note that this information is devoid of any *fear content*. The triggering of fear happens at the next level, by the appraisal processes that consciously or subconsciously decide that there is a danger in terms of the valuation described in section . The appraisal processes are modulated by the contents of the agent metalevel memory, a database of past experiences of events, agent responses and results. Eventually, this process culminates in the functional reorganization of the agent to provide the best structure to generate the most adequate behaviors —e.g. freeze, fight or run.

Depending on the concrete action, the basic functional patterns of autonomic and somatic fear are three: a) blocking and defensive immobility, b) defensive action and c) defensive escape. The implementation of fear mechanisms in the brain has been studied by many authors. The most important body of work comes from LeDoux (LeDoux, 1996) but most of his work is focused on the use of classical fear conditioning as a behavioral tool for studying emotional memories and brain activation (identifying the role of the different memory systems is central to LeDoux work). LeDoux work is centered on metalevel learning processes. Closer to the morphofunctional topic of this article, Bradley (Bradley, 2000) studied the functional and structural differences between the escape response and the freezing response and how their somatic physiology is completely different. Psychophysiological responses and somatic fear have also been studied by Hamm et al. (Hamm et al., 1997), Caioppo et al. (Cacioppo et al., 2000) or Sabatinelli et al (Sabatinelli et al., 2001).

The patterned metacontrol model of emotion addresses the dynamic reorganization of biological agents to maximize their ecological resilience. LeDoux uses the term ″survival circuit″ to refer to this very same vision: *″survival circuits are sensory-motor integrative devices that serve specific adaptive purposes″* (LeDoux, 2012).

Mapping the Theory into Machines

The approach to transfer this model to machines is quite different from what can be found in current research on artificial emotion (Bartneck et al., 2008). It is focused on the core architectural issues more than in emotion externalization. Mapping the patterned metacontrol model of emotions into machine realizations imply the construction of changeable cognitive architectures. The pattern-based approach is specially suitable for modular block-based systems. This is the best architectural primitive to support the dynamic reorganization from a pattern set of reduced cardinality. Issues of granularity may appear as finer-grained systems can offer improved flexibility but may require computationally demanding reconfiguration algorithms. The pattern-based approach, however, may help bounding this complexity (Tchao et al., 2011).

The most adequate construction and deployment strategy for this class of systems is the use of a plastic integration infrastructure that provides the required homogeneity and interface minimization. Modern software middleware platforms can help with this (esp. distributed, real-time system platforms (Sanz et al., 2001),(OMG, 2007))

The technological path to leverage this model as an universal technology of artificial emotion includes two major activities that shall be addressed:

- The identification of the organizational patterns: the definition of the enactable collection of patterns that a particular agent may use. This includes both persistent patterns and dynamic patterns (i.e. with different life-cycles). Some of these patterns may be reverse-engineered from biological/brain systems but in most cases they will be the result of classical engineering design efforts. An example of these patterns are the alternative organizations used in some fault-tolerant systems, that are enacted as needed driven by the internal detection of faults.

- The plastic engineering infrastructure: the technological substrate to sustain the functional reconfiguration has requirements that go beyond state-of-the-art real-time middleware technologies. These infrastructures shall be componential, self-referential, run-time changeable, real-time, multigranular, portable and transparent. Some of these capabilities are already available (e.g. in Real-Time CORBA) but others are still under development (Hernández et al., 2009).

We are doing an implementation of this technology in our laboratory and using it in robot control. We have completed Phase 0, developing an ad-hoc implementation of rule-based metacontrol over rigid control patterns (Bajuelos, 2011). The modular controllers used in our robot are based on the instrumented and widely available ROS robot software (Quigley et al., 2009).

The ongoing Phase 1 works are focused on the generalization of the reconfiguration methods, incorporating explicit representations of architectural and mission elements. The main aspects are related to the use of a rigorous ontology for cognitive-emotional systems (Bermejo-Alonso et al., 2010), the use of explicit representations of functions and goals (de la Mata and Rodríguez, 2010) and the use of model-based reasoning at the software system level (Hernández and Sanz, 2012). Immediate future work is related to the realization of a bus-based emotion reconfiguration architecture (Sanchez-Escribano and Sanz, 2012), the incorporation of pattern learning mechanisms and the realization of some consciousness patterns (Sanz et al., 2007b).

The main problems that are still open in the ASys implementation of the metacontrol system are related to:

- The dual representation of function in both implementational and goal terms.
- The processes of emotion dynamics — how to implement pattern change without losing dependability.
- Concurrent emotions and emotion composition — pattern coexistence, integration and interference.
- Heterogeneous agent emotions — artificial theories-of-mind (ToM), emotional societies.
- Perception of the emotional state — persistent questions of self and qualia.

Conclusions

In this article we have presented the patterned metacontrol model of emotions, where emotions are seen as processes of goal-centric metacontrol of patterned control architectures of autonomous agents.

In the biological world these patterns are created using evolutionary mechanisms. In the case of artificial systems the patterns are ad-hoc designs but can also be selected evolutionarily alongside product lines.

This model of emotion shows that the capability for pattern-based adaptation appears at two time scales: 1) at evolutionary time, by the creation of new patterns and 2) at individual time, by the activation of those that are more relevant for the present state of affairs. Additionally, in metacognitive agents patterns can also be learnt by individual agents at a cognitive level and afterwards automated.

This model of emotion can provide a theoretical framework for the explanation of biological emotion. Emotions as global synthetic signals produced by appraisal systems vehiculate reconfiguration processes of the functional organization of the agent (morphofunctional processes). These appraisal-reorganization processes occur at many levels of the agent control architecture; they occur both at the conscious and unconscious levels. At the conscious level the agent can perceive and experiment its very own emotional processes, hence having emotional qualia.

This morphofunctional model of emotion can also be implemented and leveraged in technical systems. We have described our initial steps in the process of developing cross-domain software technology for its application in complex control systems. The full potential of these approaches can only be manifest when architectural organization alternatives are available. However, current robotic architectures and infrastructures are in general too simple to fully leverage this vision.

Several theoretical issues are still open, for example: How are patterns composed? How do concurrent emotions merge? Can this model explain *all* biological emotions or is it just a model suitable for a limited class of them (e.g. the fight-or-flight response)?

It is of special relevance the question of to what extent is this model applicable. Emotion phenomena are enormously varied and this model only addresses core architectural issues. However, this core approach can sustain derived explanations of other aspects. For example, the expression of emotion —e.g. in faces or in language— can be seen as an externalization of the core functional state for social communication in search of agent benefit. Even when plenty of work awaits, we consider that this theoretical model of patterned metacontrol is a good starting point for a universal, cybernetic theory of emotion.

References

Adolphs, R., Tranel, D., Damasio, H., R.Damasio, A., 1995. Fear and the human amygdala. The Journal of Neuroscience 15, 5879–5891.

Aldridge, H., Bluethmann, B., Ambrose, R., Diftler, M., 2000. Control architecture for the robonaut space humanoid, in: First IEEE-RAS International Conference on Humanoid Robots, Cambridge, Massachusetts.

Alexander, C., Ishikawa, S., Silverstein, M., 1977. A Pattern Language: Towns, Buildings, Construction. Oxford University Press.

Anderson, J.R., Bothell, D., Byrne, M.D., 2004. An integrated theory of the mind. Psychological Review 111, 1036–1060.

Andrew, A.M., 2009. A missing link in cybernetics: logic and continuity. volume v. 26 of *International Federation for Systems Research international series on systems science and engineering*. Springer, New York.

Arbib, M.A., Fellous, J.M., 2004. Emotions: from brain to robot. Trends in Cognitive Sciences 8, 554–561.

Ashby, A.R., 1956. An Introduction to Cybernetics. Chapman and Hall, London.

Baars, B.J., Gage, N.M., 2007. Cognition, Brain and Consciousness: Introduction to Cognitive Neuroscience. Academic Press. 1 edition.

Bajuelos, A., 2011. Improving robustness in robotic navigation by using a self-reconfigurable control system. Master's thesis. Escuela Técnica Superior de Ingenieros Industriales - Universidad Politécnica de Madrid.

Balmelli, L., Brown, D., Cantor, M., Mott, M., 2006. Model-driven systems development. IBM Systems journal 45, 569–585.

Barker, F.I., 1995. Phineas among the phrenologists: the american crowbar case and nineteenth-century theories of cerebral localization. Journal of Neurosurgery 82, 672–682.

Bartneck, C., 2002. Integrating the occ model of emotions in embodied characters. Design , 1–5.

Bartneck, C., Lyons, M., Saerbeck, M., 2008. The relationship between emotion models and artificial intelligence, in: Proceedings of the Workshop on The Role Of Emotion In Adaptive Behavior And Cognitive Robotics, in affiliation with the 10th International Conference on Simulation of Adaptive Behavior: From Animals to Animates (SAB 2008), Osaka.

Bass, L., Clements, P., Kazman, R., 2013. Software Architecture in Practice. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA. third edition.

Bass, L., Cohen, S., Northrop, L., 1996. Product line architectures, in: Proceedings for the International Workshop on Development and Evolution of Software Architectures for Product Families, Las Navas del Marqués, Ávila, Spain.

Bayar, S., Yurdakul, A., 2012. A dynamically reconfigurable communication architecture for multicore embedded systems. Journal of Systems Architecture 58, 140–159.

Bermejo-Alonso, J., Sanz, R., Rodríguez, M., Hernández, C., 2010. An ontological framework for autonomous systems modelling. International Journal on Advances in Intelligent Systems 3, 211–225.

Bosse, T., Jonker, C.M., Treur, J., 2008. Formalisation of damasio's theory of emotion, feeling and core consciousness. Consciousness and Cognition In Press.

Bradley, M.M., 2000. Emotion and motivation, in: Cacioppo, J., Tassinary, L., Bernston, G. (Eds.), Handbook of Psychophysiology. Cambridge Univ. Press, New York, pp. 602– 642.

Breazeal, C., 2004. Function meets style: insights from emotion theory applied to hri. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 34, 187–194.

Brun, Y., Desmarais, R., Geihs, K., Litoiu, M., Lopes, A., Smit, M., 2013. A design space for self-adaptive systems, in: Lemos, R., Giese, H., Müller, H.A., Shaw, M. (Eds.), Software Engineering for Self-Adaptive Systems II. Springer Berlin Heidelberg. volume 7475 of *Lecture Notes in Computer Science*, pp. 33–50.

Bush, G., Luu, P., Posner, M.I., 2000. Cognitive and emotional influences in anterior cingulate cortex. Trends in cognitive sciences 4, 215–222.

Cacioppo, J., Berntson, G., Larsen, J., Poehlmann, K., Ito, T., et al., 2000. The psychophysiology of emotion. Handbook of emotions 2, 173–191.

Chaput, H.H., 2004. The Constructivist Learning Architecture: A Model of Cognitive Development for Robust Autonomous Robots. Ph.D. thesis. Computer Science Department, University of Texas at Austin. Technical Report TR-04-34.

Cheng, S.W., Garlan, D., Schmerl, B., 2006. Architecture-based self-adaptation in the presence of multiple objectives, in: ICSE 2006 Workshop on Software Engineering for Adaptive and Self-Managing Systems (SEAMS), Shanghai, China.

Cloutier, R.J., Verma, D., 2007. Applying the concept of patterns to systems architecture. Systems Engineering 10, 138–154.

Cosmides, L., Tooby, J., 2003. Evolutionary psychology: Theoretical foundations, in: Encyclopedia of Cognitive Science. Macmillan, London.

Damasio, A., 1994. Descartes' Error: Emotion, Reason and the Human Brain. Putman.

Damasio, A.R., Everitt, B.J., Bishop, D., 1996. The somatic marker hypothesis and the possible functions of the prefrontal cortex [and discussion]. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 351, 1413–1420.

Dario, P., Carrozza, M., Guglielmelli, E., Laschi, C., Menciassi, A., Micera, S., Vecchi, F., 2005. Robotics as a future and emerging technology: biomimetics, cybernetics, and neuro-robotics in european projects. Robotics Automation Magazine, IEEE 12, 29 – 45.

Davidson, R., Sutton, S., 1995. Affective neuroscience: The emergence of a discipline. Current Opinion in Neurobiology 5, 217–224.

Davidson, R.J., Lewis, D., Alloy, L., Amaral, D., Bush, G., Cohen, J., Drevets, W., Farah, M., Kagan, J., McClelland, J., Nolen-Hoeksema, S., Peterson, B., 2002. Neural and behavioral substrates of mood and mood regulation. Biological Psychiatry 52, 478–502.

Davidson, R.J., Pizzagalli, D., Nitschke, J.B., Kalin, N.H., 2003. Parsing the subcomponents of emotion and disorders of emotion: Perspectives from affective neuroscience, in: R. J. Davidson, H. H. Goldsmith, .K.S. (Ed.), Handbook of Affective Science. Oxford University Press, New York, pp. 8–24.

Deconinck, G., Vounckx, J., Cuyvers, R., Lauwereins, R., 1994. Fault tolerance in massively parallel systems. Transputer Communications 2, 241–257.

Dennett, D.C., 1996. Kinds of minds: toward an understanding of consciousness. Science masters, Basic Books, New York, NY. 1st ed edition.

Dolcos, F., Iordan, A.D., Dolcos, S., 2011. Neural correlates of emotion—-cognition interactions: A review of evidence from brain imaging investigations. Journal of Cognitive Psychology 23, 669–694.

Dyer, M.G., 1987. Emotions and their computations: Three computer models. Cognition & Emotion 1, 323–347.

Edwards, C., Spurgeon, S., 1998. Sliding Mode Control: Theory and Applications. Taylor and Francis, London.

Eibl-Eibesfeldt, I., 1971. Love and hate: on the natural history of basic behaviour patterns. Methuen, London.

Famaey, J., Latré, S., Strassner, J., De Turck, F., 2010. An ontology-driven semantic bus for autonomic communication elements, in: Brennan, R., Fleck, J., van der Meer, S. (Eds.), Modelling Autonomic Communication Environments. Springer Berlin / Heidelberg. volume 6473 of *Lecture Notes in Computer Science*, pp. 37–50.

Fellous, J.M., 2004. From human emotions to robot emotions, in: AAAI Spring Symposium on Architectures for Modeling Emotion, Stanford University, USA.

Fellous, J.M., Arbib, M.A., 2005. Who needs emotions?: the brain meets the robot. Oxford University Press, Oxford.

Fiadeiro, J.L., Lopes, A., 2010. A model for dynamic reconfiguration in service-oriented architectures, in: Proceedings of the 4th European conference on Software architecture, Springer-Verlag, Berlin, Heidelberg. pp. 70–85.

Frijda, N.H., 1986. The Emotions. Cambridge University Press.

Frijda, N.H., 2007. The laws of emotion. Lawrence Erlbaum Associates, Mahwah, N.J.

Gamma, E., Helm, R., Johnson, R., Vlissides, J., 1995. Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley Professional Computing Series, Addison-Wesley Publishing Company, New York, NY.

Gini, G.C., Folgheraiter, M., Scarfogliero, U., Moro, F., 2009. A biologically founded design and control of a humanoid biped, in: Choi, B. (Ed.), Humanoid Robots. InTech. chapter 3.

Gleeson, P., Crook, S., Cannon, R.C., Hines, M.L., Billings, G.O., Farinella, M., Morse, T.M., Davison, A.P., Ray, S., Bhalla, U.S., Barnes, S.R., Dimitrova, Y.D., Silver, R.A., 2010. Neuroml: A language for describing data driven models of neurons and networks with a high degree of biological detail. PLoS Comput Biol 6, e1000815.

Gómez, J., Sanz, R., Hernández, C., 2008. Cognitive ontologies: Mapping structure and function of the brain from a systemic view, in: AAAI 2008 Fall Symposium on Biologically Inspired Cognitive Architectures.

Gould, S.J., 2002. The structure of evolutionary theory. Belknap Press of Harvard University Press, Cambridge, Mass.

Hamm, A.O., Cuthbert, B.N., Globisch, J., Vaitl, D., 1997. Fear and the startle reflex: Blink modulation and autonomic response patterns in animal and mutilation fearful subjects. Psychophysiology 34, 97–107.

Hansen, L.P., Sargent, T.J., 2007. Robustness. Princeton University Press.

Hara, F., Pfeifer, R., 2003. Morpho-functional machines: the new species : designing embodied intelligence. Springer, Berlin.

Hernández, C., López, I., Sanz, R., 2009. The operative mind: a functional, computational and modelling approach to machine consciousness. International Journal of Machine Consciousness 1, 83–98.

Hernández, C., Sanz, R., 2012. Three Patterns for Autonomous Systems. Technical Note ASLAB-A-2012-007.pdf. Autonomous Systems Laboratory ASLab. Universidad Politécnica de Madrid.

Hernández, C., Sanz, R., Gómez-Ramirez, J., Smith, L.S., Hussain, A., Chella, A., Aleksander, I. (Eds.), 2011. From Brains to Systems. Brain-Inspired Cognitive Systems 2010. Springer.

Hernandez, C., Sanz, R., Hernando, A., Gomez, J., Sanchez, G., Salom, M., 2010. Integrated Cognitive Architecture Model. Deliverable D61. Universidad Politécnica de Madrid - European Integrated Project IST-027819 ICEA.

Herrera, C., Sánchez, G., Sanz, R., 2012. The morphofunctional approach to emotion modelling in robotics. Adaptive Behavior 20, 388–404.

Herrera-Perez, C., Sanz, R., 2013. Emotion as morphofunctionality. Artificial Life 19.

Houwer, J.D., Hermans, D., 2010. Cognition & Emotion: Reviews of Current Research and Theories. Psychology Press.

Klir, G.C., 1969. An Approach to General Systems Theory. Van Nostrand Reinhold.

Kluwer, H., Bucy, P.C., 1939. Preliminary analysis of functions of the temporal lobe in monkeys. Archives of Neurology and Psychiatry 42, 979–1000.

Krichmar, J.L., 2012. Design principles for biologically inspired cognitive robotics. Biologically Inspired Cognitive Architectures 1, 73–81.

Kringelbach, M.L., Berridge, K.C., 2009. Towards a functional neuroanatomy of pleasure and happiness. Trends in Cognitive Sciences 13, 479 – 487.

Kushiro, K., Harada, Y., Takeno, J., 2013. Robot uses emotions to detect and learn the unknown. Biologically Inspired Cognitive Architectures This Issue.

Laird, J.E., Newell, A., Rosenbloom, P.S., 1987. Soar: an architecture for general intelligence. Artificial Intelligence 33, 1–64.

Laird, P.R.J., Newell, A. (Eds.), 1993. The Soar Papers: Research on Integrated Intelligence. MIT Press, Cambridge, MA.

van Lamsweerde, A., 2003. From system goals to software architecture, in: Bernardo, M., Inverardi, P. (Eds.), Formal Methods for Software Architecture. Springer.

Larue, O., Poirier, P., Nkambou, R., 2013. The emergence of (artificial) emotions from cognitive and neurological processes. Biologically Inspired Cognitive Architectures This Issue.

Lazarus, R.S., 1991. Emotion and adaptation. Oxford University Press, New York.

LeDoux, J., 1996. The Emotional Brain. Simon & Schuster, New York.

LeDoux, J., 2012. Rethinking the emotional brain. Neuron 73, 653–676.

Lind, M., 1990. Representing Goals and Functions of Complex Systems. An Introduction to Multilevel Flow Modelling. Technical Report. Institute of Automatic Control Systems. Technical University of Denmark.

Lindquist, K.A., Barrett, L.F., 2012. A functional architecture of the human brain: emerging insights from the science of emotion. Trends in Cognitive Sciences 16.

Lipson, H., Pollack, J.B., 2000. Automatic design and manufacture of robotic lifeforms. Nature 406, 974–980.

Liu, J.W., 2000. Real-time Systems. Prentice-Hall, Upper Saddle River, NJ.

Lorini, E., 2011. Qualitative and quantitative aspects of expectation-based emotions: a logical approach, in: Broekens, J., Bosse, T., Marsella, S. (Eds.), Workshop on Standards in Emotion Modeling, Lorentz Center.

MacLean, P.D., 1949. Psychosomatic disease and the 'visceral brain': recent developments bearing on the papez theory of emotion. Psychosom. Med. 11, 38 –353.

Marinier III, R.P., Laird, J.E., Lewis, R.L., 2009. A computational unification of cognitive behavior and emotion. Cognitive Systems Research 10, 48–69.

Marsella, S., Gratch, J., Petta, P., 2010. Computational models of emotion, in: Scherer, K.R., Banziger, T., Roesch, E. (Eds.), A Blueprint for Affective Computing: A sourcebook and manual. Oxford University Press, USA, Oxford, UK. Series in affective science.

de la Mata, J.L., Rodríguez, M., 2010. Accident prevention by control system reconfiguration. Computers and Chemical Engineering 34, 846–855.

McClelland, J.L., 2009. The place of modeling in cognitive science. Topics in Cognitive Science 1, 11 – 38.

Miller, E., Cohen, J., 2001. An integrative theory of prefrontal cortex function. Annu Rev Neurosci. 24, 167–202.

Mitchinson, B., Chan, T., Chambers, J., Humphries, M., Gurney, K., Prescott, T., 2008. Brahms: Novel middleware for integrated systems computation, in: Frontiers in Neuroinformatics. Conference Abstract: Neuroinformatics 2008.

Moffat, D., Frijda, N., Phaf, H., 1995. Analysis of a model of emotions, in: Sloman, A., Hogg, D., Humphreys, G., Ramsay, A., Partridge, D. (Eds.), Prospects for Artificial Intelligence: Proceedings of AISB93.

Morris, J.S., Ohman, A., Dolan, R.J., 1998. Conscious and uncon- scious emotional learning in the human amygdala. Nature 393, 467– 470.

Nolfi, S., Floreano, D., 2000. Evolutionary robotics: the biology, intelligence, and technology of self-organizing machines. MIT Press, Cambridge, Mass.

Ochsner, K.N., Bunge, S.A., Gross, J.J., Gabrieli, J.D.E., 2002. Rethinking feelings: An fMRI study of the cognitive regulation of emotion. Journal of Cognitive Neuroscience 14, 1215–1299.

OMG, 2007. Data Distribution Service for Real-time Systems. OMG Available Specification formal/07-01-01. Object Management Group.

Panksepp, J., 1982. Toward a general psychobiological theory of emotions. Behavioral and Brain Sciences 5, 407–422.

Panksepp, J., 1998. Affective neuroscience: the foundations of human and animal emotions. Oxford University Press, New York.

Papez, J.W., 1995. A proposed mechanism of emotion. 1937. The Journal of Neuropsychiatry and Clinical Neurosciences 7, 103–112.

Phelps, E., 2006. Emotion and cognition: Insights from studies of the human amygdala. Annual Review of Psychology 24, 27–53.

Phelps, E., LeDoux, J., 2005. Contributions of the amygdala to emotion processing: From animal models to human behavior. Neuron 48, 175–187.

Picard, R.W., 1997. Affective Computing. The MIT Press.

Quigley, M., Conley, K., Gerkey, B.P., Faust, J., Foote, T., Leibs, J., Wheeler, R., Ng, A.Y., 2009. Ros: an open-source robot operating system, in: ICRA Workshop on Open Source Software.

Ritter, F.E., 1993. Three types of emotional effects that will occur in a cognitive architecture like SOAR, in: Workshop on architectures underlying motivation and emotion, University of Birmingham.

Rolls, E.T., 2005. Emotion explained. Oxford University Press, Oxford.

Rosen, J.B., Donley, M.P., 2006. Animal studies of amygdala function in fear and uncertainty: Relevance to human research. Biological Psychology 73, 49–60.

Rosen, R., 1970. Structural And Functional Considerations In The Modelling Of Biological Organization. Report 77-25. Center for the Study of Democratic Institutions.

Rosen, R., 2012. Anticipatory Systems. Philosophical, Mathematical, and Methodological Foundations. volume 1 of *IFSR International Series on Systems Science and Engineering*. Springer. 2nd edition.

Rosenthal, D.M., 2009. Higher-order theories of consciousness, in: McLaughlin, B., Beckermann, A., Walter, S. (Eds.), Oxford Handbook on the Philosophy of Mind. Oxford University Press, Oxford. chapter 13.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536.

Sabatinelli, D., Bradley, M.M., Lang, P.J., 2001. Affective startle modulation in anticipation and perception. Psychophysiology 38, 719–72.

Samsonovich, A.V., 2012. On a roadmap for the bica challenge. Biologically Inspired Cognitive Architectures 1, 100–107.

Sanchez-Escribano, M.G., Sanz, R., 2012. Value by architectural transversality, emotion and consciousness, in: 16th Meeting of the Association for the Scientific Study of Consciousness, Brighton, UK.

Sanz, R., 2003. Against biologism. contribution to the panel on intelligent control imitating biology: Promises, challenges and lessons., in: IEEE International Conference on Intelligent Control, IEEE Press, Houston, USA.

Sanz, R., Alonso, M., López, I., García, C.A., 2001. Enhancing robot control architectures using CORBA, in: Proceedings of IEEE International Symposium on Intelligent Control, Mexico D.F.

Sanz, R., Hernández, C., Gómez, J., Bedia, M.G., 2010a. Against animats, in: Proceedings of CogSys 2010 - 4th International Conference on Cognitive Systems, Zurich, Switzerland.

Sanz, R., Hernández, C., Gómez, J., Bermejo-Alonso, J., Rodríguez, M., Hernando, A., Sánchez, G., 2009. Systems, models and self-awareness: Towards architectural models of consciousness. International Journal Of Machine Consciousness 1, 255–279.

Sanz, R., Hernández, C., Gómez, J., Sánchez, G., Hernando, A., 2010b. About the validity of computer models in cognitive science, in: Proceedings of the 32nd Annual Meeting of the Cognitive Science Society, Portland, USA.

Sanz, R., Hernández, C., Rodriguez, M., 2010c. The epistemic control loop, in: Proceedings of CogSys 2010 - 4th International Conference on Cognitive Systems, Zurich, Switzerland.

Sanz, R., Hernandez, C., Sanchez-Escribano, M.G., 2012. Consciousness, action selection, meaning and phenomenic anticipation. International Journal of Machine Consciousness 4, 383–399.

Sanz, R., López, I., 2000. Minds, MIPS and structural feedback, in: Performance Metrics for Intelligent Systems, PerMIS '2000, Gaithersburg, USA.

Sanz, R., López, I., Bermejo-Alonso, J., 2007a. A rationale and vision for machine consciousness in complex controllers, in: Chella, A., Manzotti, R. (Eds.), Artificial Consciousness. Imprint Academic, pp. 141–155.

Sanz, R., Lopez, I., Hernandez, C., Gomez, J., 2008. Emotion in the ASys framework, in: Brain-Inspired Cognitive Systems Conference, Sao Luis, Brazil.

Sanz, R., López, I., Rodríguez, M., Hernández, C., 2007b. Principles for consciousness in integrated cognitive control. Neural Networks 20, 938–946.

Sanz, R., Matía, F., Galán, S., 2000. Fridges, elephants and the meaning of autonomy and intelligence, in: IEEE International Symposium on Intelligent Control, ISIC'2000, Patras, Greece.

Scherer, K.R., 2000. Emotions as episodes of subsystem synchronization driven by nonlinear appraisal processes, in: (Eds.), M.D.L..I.G. (Ed.), Emotion, development, and self-organization:Dynamic systems approaches to emotional development. Cambridge University Press, New York / Cambridge, pp. 70–99.

Scherer, K.R., 2009. Emotions are emergent processes: they require a dynamic computational architecture. Philosophical Transactions of the Royal Society B: Biological Sciences 364, 3459–3474.

Scherer, K.R., Banziger, T., Roesch, E.B., 2010. Blueprint for affective computing: a sourcebook and manual. Series in affective science, Oxford University Press, Oxford.

Scherer, K.R., Shorr, A., Johnstone, T. (Eds.), 2001. Appraisal processes in emotion: theory, methods, research. Oxford University Press, Canary, NC.

Selfridge, O.G., 1958. Pandemonium: A paradigm for learning, in: Blake, D.V., Uttley, A.M. (Eds.), Proceedings of the Symposium on Mechanisation of Thought Processes, National Physical Laboratory. Her Majesty's Stationery Office, London. pp. 511–529.

Sellers, M., 2013. Toward a general theory of emotion for biological and artificial agents. Biologically Inspired Cognitive Architectures This Issue.

Serugendo, G.D.M., Gleizes, M.P., Karageorgos, A., 2011. Self-organising software: from natural to artificial adaptation. Natural computing series, Springer Berlin Heidelberg, New York, NY.

Shackman, A.J., Salomons, T.V., Slagter, H.A., Fox, A.S., Winter, J.J., Davidson, R.J., 2011. The integration of negative affect, pain, and cognitive control in the cingulate cortex. Nature Reviews Neuroscience 12, 154–167.

Shaw, M., Clements, P.C., 1997. A field guide to boxology: Preliminary classification of architectural styles for software systems, in: Proceedings of the 21st International Computer Software and Applications Conference, IEEE Computer Society, Washington, DC, USA. pp. 6–13.

Simon, H.A., 1996. The Sciences of the Artificial. MIT Press, Cambridge,USA. third edition.

Sloman, A., 1995. Exploring design space and niche space, in: Proceedings 5th Scandinavian Conf. on AI, Trondheim, IOS Press, Amsterdam.

Stewart, T.C., West, R.L., 2007. Deconstructing and reconstructing ACT-R: Exploring the architectural space. Cognitive Systems Research 8, 227 – 236. Cognitive Modeling.

Sun, R. (Ed.), 2008. The Cambridge Handbook of Computational Psychology. Cambridge University Press, New York.

Sztipanovits, J., Karsai, G., 1997. Model-integrated computing. Computer 30, 110–111.

Tamburrini, G., Datteri, E., 2005. Machine experiments and theoretical modelling: from cybernetic methodology to neuro-robotics. Minds and Machines 15, 335–358.

Tchao, A.E., Risoldi, M., Di Marzo Serugendo, G., 2011. Modeling self-* systems using chemically-inspired composable patterns, in: Self-Adaptive and Self-Organizing Systems (SASO), 2011 Fifth IEEE International Conference on, pp. 109 –118.

Thórisson, K.R., 2009. From constructionist to constructivist A.I., in: AAAI Fall Symposium on Biologically-Inspired Cognitive Architectures, AAAI press, Washington DC, USA. pp. 175–183.

Treur, J., 2013. An integrative dynamical systems perspective on emotions. Biologically Inspired Cognitive Architectures This Issue.

Webb, B., 2009. Animals versus animats: Or why not model the real iguana? Adaptive Behavior 17, 269–286.

Webb, B., Consi, T.R. (Eds.), 2001. Biorobotics. Methods and Applications. The MIT Press, Cambridge, MA, USA.

Weston, R.H., 1999. Reconfigurable, component-based systems and the role of enterprise engineering concepts. Computers in Industry 40, 321–343.

Wilson-Mendenhall, C.D., Barrett, L.F., Simmons, W.K., Barsalou, L.W., 2011. Grounding emotion in situated conceptualization. Neuropsychologia 49, 1105–1127.

Ziemke, T., 2008. On the role of emotion in biological and robotic autonomy. Biosystems 91, 401–408.

Zinober, A.S. (Ed.), 1994. Variable Structure and Lyapunov Control. Springer-Verlag, London.