

# HUMAINE

**D7b**

**Preliminary plans for exemplars:  
Cognition & Action**

**Lola Cañamero and WP7 members**



**Version 1.0**

**Date: 28 May 2004**

<b>IST project contract no.</b>	507422
<b>Project title</b>	<b>HUMAINE Human-Machine Interaction Network on Emotions</b>
<b>Contractual date of delivery</b>	<i>month 5</i>
<b>Actual date of delivery</b>	<i>28 May 2004</i>
<b>Deliverable number</b>	<i>D7b</i>
<b>Deliverable title</b>	Preliminary plans for exemplars: Cognition & Action
<b>Type</b>	Public Report
<b>Number of pages</b>	57
<b>WP contributing to the deliverable</b>	WP7
<b>Task responsible</b>	UH
<b>Author(s)</b>	Lola Cañamero and WP7 members
<b>EC Project Officer</b>	Philippe Gelin

Address of author:

Lola Cañamero  
*School of Computer Science  
 University of Hertfordshire  
 College Lane  
 Hatfield, Herts AL10 9AB  
 United Kingdom*

## Table of Contents

<b>1</b>	<b>THE STATUS OF THIS REPORT .....</b>	<b>5</b>
<b>2</b>	<b>THEMATIC DEFINITION OF THIS WORKPACKAGE.....</b>	<b>6</b>
2.1	The role of WP 7 within HUMAINE.....	6
2.2	Key issues.....	8
<b>3</b>	<b>REVIEW OF KEY CONCEPTS IN THE THEMATIC AREA .....</b>	<b>10</b>
<b>4</b>	<b>REVIEW OF KEY ACHIEVEMENTS IN THE THEMATIC AREA.....</b>	<b>17</b>
4.1	Emotion-based architectures for adaptive autonomous agents.....	17
4.2	Appraisal systems .....	20
4.3	User modeling.....	21
4.4	Embodied conversational agents and virtual environments.....	22
<b>5</b>	<b>REVIEW OF KEY PROBLEMS IN THE THEMATIC AREA.....</b>	<b>24</b>
5.1	Conceptual problems.....	24
5.2	Integration Challenges .....	27
<b>6</b>	<b>ASSESSMENT OF THE KEY DEVELOPMENT GOALS IN THE THEMATIC AREA .....</b>	<b>30</b>
6.1	Some current “bottlenecks” that need to be addressed.....	30
6.2	Key development goals .....	31
<b>7</b>	<b>RELATION TO OTHER WORKPACKAGES .....</b>	<b>34</b>
7.1	WP 3: Theories and models .....	34
7.2	WP 4: Signals to signs of emotions.....	34
7.3	WP 5: Data and databases .....	35

---

<b>7.4</b>	<b>WP 6: Emotion in interaction .....</b>	<b>35</b>
<b>7.5</b>	<b>WP 8: Emotion in communication and persuasion .....</b>	<b>36</b>
<b>7.6</b>	<b>WP 9: Usability .....</b>	<b>36</b>
<b>7.7</b>	<b>WP 10: Ethics and good practice .....</b>	<b>37</b>
<b>8</b>	<b>PRELIMINARY IDEAS ABOUT POSSIBLE EXEMPLARS .....</b>	<b>38</b>
<b>8.1</b>	<b>First Possible Exemplar: Emotion in “lower-level” cognition and action .....</b>	<b>38</b>
<b>8.2</b>	<b>Second Possible Exemplar: Emotion in “higher-level” cognition and action .....</b>	<b>41</b>
<b>8.3</b>	<b>Third Possible Exemplar: bridging the gap between “lower-level” and “higher-level” cognition and action .....</b>	<b>42</b>
<b>8.4</b>	<b>Fourth Possible Exemplar: Emotions in Social Cognition and Interaction .....</b>	<b>44</b>
<b>9</b>	<b>CONCLUSIONS AND WAY FORWARD .....</b>	<b>48</b>
	<b>REFERENCES .....</b>	<b>49</b>
	<b>References cited in the text .....</b>	<b>49</b>
	<b>General bibliography.....</b>	<b>56</b>

## 1 The status of this report

Joint research in HUMAINE aims to produce 'exemplars'. We chose the term exemplars to convey that above all, the systems that we develop should embody sound principles. The systems may be working models or 'in principle' specifications. Embodying sound principles means not only that they should exemplify good ways of addressing individual problems, but also that the set of exemplars taken as a whole defines a rational ways of partitioning the overall problem of developing emotion-sensitive systems. Arriving at a satisfying partition is a major part of the challenge that HUMAINE faces, requiring iteration and consultation between groups dealing with different thematic areas.

The Technical Annex sets process that is designed to meet that challenge. It will begin with production by each thematic group of a review of key concepts, achievements and problems in its thematic area; and drawn from the review, an assessment of the key development goals in the area. This review and assessment will be circulated to the whole network for discussion and comment, aimed both at building understanding of basic issues across areas, and at identifying the choices of goal that would be most likely to let the different groups achieve complementary developments. That consultation phase will provide the basis for deliverables in month 11, which will specify a range of exemplars that deserve serious consideration. A further round of consultation will follow before concrete plans for each workpackage are drawn up and shared at the 18-month plenary meeting.

This report is the review that defines the starting point of the process for work package WP7, whose brief title is 'Emotion in Cognition and Action'.

## 2 Thematic definition of this workpackage

In humans, emotions entail distinctive integrated ways of perceiving and assessing situations, processing information, and modulating and prioritizing actions. Elaborating computational models that embed the effects of emotions in cognition and action is a complex, multi-faceted problem that poses multiple conceptual, technical and integration challenges. WP 7 aims to lay the foundations for addressing these issues in a sound, coherent and integrated way. Its goal is thus to achieve a better understanding of basic issues, key developments, open research topics, and key application scenarios regarding the involvement of emotions in cognition and action, with a view to grounding and promoting sound research into artificial emotional systems for artifacts that must interact with humans.

This goal necessitates an integrated effort spanning different disciplines, rather than the development of isolated engineering projects. A shared critical reflection in this area is needed regarding, among other aspects: key conceptual issues, assumptions and dependencies; open research problems; key research and application scenarios; analysis of needs and directions for future research and applications based on a critical analysis of existing approaches, systems and tools; evaluation methods, scenarios and tools; analysis of needs for recommendations of good practice; requirements for usability; and potential for cross-fertilization among disciplines.

It is expected that an integrative effort in this area would shed light towards the development of sound computational models of emotions that at the same time (a) enhance the behavior of emotion-oriented systems and our interactions with them, and (b) provide feedback to emotion theorists to gain further insight in their understanding of human emotions. In this respect, the contribution of such models is twofold. On the one hand, they endow the observable behavior of the artifact with the features of autonomy and coherence that are required to achieve long-term interactions adapted to humans. On the other hand, they contribute towards a better understanding of human emotions by providing a synthetic approach (by building systems) that complements the analytic studies carried out in disciplines such as psychology and cognitive neuroscience.

### 2.1 The role of WP 7 within HUMAINE

The interpersonal nature of human cognition is an important feature that must be taken into account when designing emotion-oriented systems for human-machine interaction: to be adapted to (meaningful to and accepted by) the human side of the interaction ‘loop,’ such systems must be designed to make them appear as ‘life-like,’ ‘believable’ entities and social partners to humans. Emotions and their expression are one of the key factors influencing human perception and attribution of ‘life-like’ properties to other entities, both biological and artificial (Reeves and Nass, 1996), properties that make perceive such entities as being ‘like us’ and therefore, believable social partners. A first step towards the goal of achieving ‘life-like,’ ‘believable’ emotion-oriented systems is thus to endow these systems with the capability to generate and respond to the external manifestations of emotions. Within HUMAINE, these aspects of perception, interpretation, generation of and social response to the external manifestations of emotions are investigated under other Workpackages, namely: WP 4, From Signals to Emotional Signs; WP 6, Emotion in Interaction; WP8, Emotion in Communication and Persuasion; and WP 9, Usability.

However, in our opinion, this goal of achieving believable and user-adapted emotion-oriented systems cannot be properly achieved by modeling (and responding to) only the externally observable features of emotional expression and behavior; on the contrary, it requires grounding these externally observable features in models of the (internal) mechanisms underlying the involvement of emotions in different aspects of cognition and in the production of observable emotional manifestations. This viewpoint therefore takes a position in the debate between “shallow” versus “deep” approaches to modeling emotions in interactive artifacts (see Cañamero, 2001a for a discussion of this debate), in their attempt to answer the question ‘what makes emotional artifacts believable for social interaction?’ In some application areas, such as art installations or the video and game industries, “shallow” modeling of the observable, surface aspects of emotions has been the most commonly favoured approach, giving rise to very successful systems (e.g., Bates, 1994; Stern, 2003). For some types of interactive systems, such an approach can therefore be sufficient and, in addition to producing believable characters, it can also provide very valuable insights regarding human perception of emotions (e.g., identifying key elements that trigger in humans the tendency to anthropomorphize the entities we interact with and what makes emotional displays and behavior believable to human eyes) and guidelines for the design of interactive emotion-oriented artifacts that look more natural to humans. However, “shallow” modeling makes it very difficult to maintain believability in the case of longer-term interactions with humans, as it is unlikely that the behavior of the artifact will be autonomous enough and remain coherent and adapted to the human over sustained periods of time. To achieve this, expressive behavior must be guided by a “deep” model, an underlying emotional system fully integrated in the architecture of the artifact. This is known as *emotion synthesis* in the affective computing literature (Picard, 1997): endowing artifacts with “deep” or “internal” mechanisms that give them the ability to generate or synthesize emotions and, in some sense, to “have” emotions. Emotion synthesis makes four major contributions to the emotion-oriented systems that are the object of HUMAINE:

- 1) An underlying emotion system increases the believability of the artifact by providing coherence to its behavior and interactions with humans, as they respond to a common and well-defined model.
- 2) Mapping the role that emotions play in the coordination and synchronization of other (cognitive, behavioral and bodily) subsystems in humans and animals, a properly grounded emotion system should play a major role in the synchronization and coordination of the overall architecture and behavior of the emotion-oriented artifact, improving its coherence not only from the point of view of the human user (believability of its observable manifestations) but also regarding the functioning and performance of the system.
- 3) When the underlying emotion system is elaborated taking inspiration from human emotional systems (or rather some aspects of them), it can serve as a “model” of the human user that can help to build an emotional and personality profile of the human user – a part of the ‘user model’.
- 4) Artificial emotion systems that are psychologically or biologically plausible can be used as “virtual laboratories” or tools that can contribute to the understanding of human emotions, providing feedback to emotion theorists with a synthetic perspective (by building systems with parameters that are easy to manipulate and test) that complements their analytic studies.

This Workpackage is thus primarily concerned with the investigation of (computational) “internal” mechanisms (emotion architectures) that allow to synthesize or generate emotions and model their involvement in various aspects of cognition and action in emotion-oriented systems.

## 2.2 Key issues

The study of emotion architectures that model and operationalize (different aspects of) the involvement of emotions in cognition and action in artifacts carries many parallels with the problems investigated by emotion theories and models (WP 3). As pointed above, however, WP 7 investigates these problems from a synthetic perspective, and this involves additional issues regarding the adequacy and integration of different AI approaches, representations and tools to model, operationalize and implement theoretical notions and models.

At a **conceptual level**, topics that can be meaningfully considered within this WP, not only regarding human emotions but also their meaning (or lack of it) and the underlying mechanisms required to obtain functional counterparts in artificial systems, include the following:

- Embodiment aspects of emotions and their influence on (lower- and higher-level) cognition and action.
- Emotions as cognitive modes and value systems to apprehend and evaluate the (physical and social) world.
- Emotional memory and learning.
- Emotional modulation of lower-level cognition: e.g., of perception, attention, etc.
- Emotional modulation of higher-level cognition: the role of emotions in decision-making, planning, problem-solving, etc.
- Mechanisms for appraisal and their neurobiological basis.
- Regulatory aspects of emotions: bodily regulation, regulation of our relationships with the external world, coping and regulation of our own and others’ emotions.
- Emotions as mechanisms for adaptation to challenging and changing environmental situations.
- The interplay between emotion and motivation in the production of action, and emotional modulation of behavior.
- The integrative roles of emotions: in coordinating bodily, cognitive and behavioral sub-systems, in the development and understanding of the concept of ‘self’, etc.
- Emotional disorders and cognitive and behavioral disorders related to or affected by emotions.

Regarding **computational modeling aspects**, some of the key issues that need to be addressed include:

- Identification of elements and features of emotions occurring in biological systems relevant to different computational models of emotions.
- Understanding of the scope, adequacy, and potential for cooperation/integration of different conceptual traditions and modeling paradigms.
- Understanding of the scope, adequacy, and potential for cooperation/integration of different representation formalisms.
- Understanding of the scope, limitations, and potential for cooperation/integration of existing emotion-based architectures and tools.

Finally, as pointed out at the beginning of this section, the goal of this WP necessitates an integrated effort spanning different disciplines. This effort should therefore entail a **philosophical reflection** that takes a broader view on, and puts in perspective, the problems underlying the involvement of emotions in cognition and action regarding, in particular:

- The relation of the problems addressed by theoretical and computational models of emotions to those addressed by different philosophical traditions (e.g., commonalities, differences).
- The potential contributions of computational models to our understanding of (human) emotions; e.g., do these models pose particular challenges? Do they propose viewpoints, methods or solutions that allow to re-conceptualize ill-posed or poorly understood questions?
- The epistemological and ethical implications of emotion synthesis (systems that “have” emotions).

We believe that such philosophical reflection can provide a very valuable contribution towards achieving a better understanding of key conceptual issues, assumptions and dependencies, and to set the ground for a sound and long-lasting cross-fertilization among the different disciplines involved in this thematic area.

To conclude this section, it is worth noting that it is NOT the aim of this Workpackage to come up with (the design of) a single emotion architecture that could be considered definitive or “best”, nor to privilege a model or paradigm over the rest. We believe that such a goal would be a misleading one, not only because the current state of the art in emotion theory and computational models makes it utterly unrealistic, but primarily because much can be gained by exploring the diversity of existing and potential computational models, their theoretical underpinnings, and application avenues. Time and resources available within the project impose obvious constraints to this task, as we will be able to address only a few of the relevant issues. However, the exemplars developed within this WP will be targeted at exploring and exploiting the richness of such diversity.

### 3 Review of key concepts in the thematic area

This section characterizes some of the main concepts use in the area of computational models of emotion in cognition and action, paying particular attention to notions relevant to the definition of exemplars. For the reader's convenience, the section is organized as a glossary. Definitions of the terms "emotion", "cognition", "action", and "computational model" are deliberately excluded, since attempts at characterizing them would be either too partial, reflecting the views of one of the many relevant theoretical and computational traditions, or too long, and therefore unsuitable as "definitions." The reader can refer to deliverable D3c for a characterization of "emotion", and understand the other terms in their intuitive sense, or extrapolate from the topics covered in the deliverable to grasp their more technical meanings.

#### **Action (or behavior) selection (versus decision making)**

The problem of action (or behavior) selection for an autonomous agent consists in making a decision as to what behavior to execute next in order to fulfill several time-dependent, conflicting goals. It opposes to the more analytic, functional, "high-level" decision-making problem, which optimizes the behavioral choice using mathematical modeling of both agent and environment. An action selection mechanism provides a "low-level" arbitration between behavioral alternatives, following the synthetic approach to artificial intelligent of "Behavior-Based Robotics" and "Embodied Artificial Intelligence".

#### **Appraisal**

Magda Arnold introduced the term appraisal in the 1960s, in the sense of direct, immediate, and intuitive evaluations, to account for qualitative distinctions among emotions. "Appraisal is the process triggered by an eliciting event wherein the subjective potential or actual significance of an event or situation is assessed; i.e., with respect to the subject's own goals, needs, and concerns on the one hand, and the capacity to adapt on the other hand." (Kappas, 2001, p.157).

A controversy about whether cognition is involved in appraisal can be considered mostly settled, along the lines cautiously put down in (Frijda, 1993, p. 379): "Then, how should one conceive of the basic processes of emotion elicitation? First, it must be admitted that it hinges upon a noncognitive step... primary appraisal often involves elaborate steps of inference and the intervention of knowledge. ...This most basic appraisal process may perhaps not meaningfully be called cognitive, as it may not always involve comparison between two representations, which might be taken as the minimal attribute of "cognition". Still, it involves some "computation" (LeDoux 1989, p. 271) and an appraisal process thus is a necessary condition for emotional experience and major aspects of emotional response.

#### **Architecture**

"The main goal of research in autonomous agents is to understand better the principles and organizations that underlie adaptive, robust, effective behavior. A secondary goal is to also develop tools, techniques, and algorithms for constructing autonomous agents that embody these principles and organizations. We call the totality of a set of principles and organizations, and the set of tools, algorithms and techniques that support them an "architecture" for modeling autonomous agents." (Maes, 1995, page 138)

Architectures operationalized in robots are often called “controllers”.

In the context of Action Selection, an (action selection) architecture specifies the way in which different architectural elements, such as internal and external stimuli, motivations, emotions, behaviors, etc. are combined to produce the final selection of one behavioral alternative.

### **Artificial Intelligence (AI)**

“Broadly (and somewhat circularly) defined, is concerned with intelligent behavior in artifacts. Intelligent behavior, in turn, involves perception, reasoning, learning, communicating, and acting in complex environments. AI has as one of its long-term goals the development of machines that can do these things as well as humans can, or probably even better. Another goal of AI is to understand this kind of behavior whether it occurs in machines or in humans or other machines.” (Nilsson, 1998)

### **Autonomous (and adaptive) agent**

“An *agent* is a system that tries to fulfill a set of goals in a complex, dynamic environment. An agent is situated in the environment: It can sense the environment through its sensors and act upon the environment using its actuators. An agent’s goals can take many different forms: They can be “end goals” or particular states the agent tries to achieve, they can be a selective reinforcement or reward that the agent attempts to maximize, they can be internal needs or motivations that the agent has to keep within certain viability zones, and so on. An agent is called *autonomous* if it operates completely autonomously, that is, if it decides itself how to relate its sensor data to motor commands in such a way that its goals are attended to successfully. An agent is said to be *adaptive* if it is able to improve over time, that is, if the agent becomes better at achieving its goals with experience. Notice that there is a continuum of ways in which an agent can be adaptive, from being able to adapt flexibly to short-term, smaller changes in the environment, to dealing with more significant and long-term (lasting) changes in the environment, that is, being able to change and improve behavior over time.” (Maes, 1995, page 136)

### **Behavior-Based Robotics**

A subdiscipline of (embodied) AI and autonomous robotics that conceives robots architectures in terms of “behaviors” or competence modules implementing the various activities that a robot can perform in the particular environment that it inhabits. A behavior-based robot has a set of behavior modules that compete with one another in order to gain control of the robot’s actuators. This discipline was born during mid 80’s as a response to the apparent “failure” of the more traditional “knowledge-based” or “top-down” Artificial Intelligence (AI) in building intelligent autonomous robots. It uses a “bottom-up” methodology to synthesize systems incrementally adding behavioral modules. It closely relates to “Embodied Artificial Intelligence”. (cf. Arkin, 1998)

### **Belief-Desire-Intention (BDI) Architecture**

Within the research community concerned with software agents, the term beliefs-desires-intentions (BDI) has been used variously to denote a position on theoretically useful mental

state distinctions, particular models of how these mental states affect reasoning and a genre of architectures or frameworks for developing software agents.

“BDI agents are rational agents having certain mental attitudes of Belief, Desire and Intention, representing respectively, the information, motivational and Deliberative states of agent. These mental attitudes determine the agent's behavior and are critical for achieving adequate or optimal performance when deliberation is subject to resource bounds.” (Rao and Georgeff, 1995).

### **Concern**

“A concern is a disposition to desire occurrence or non-occurrence of a given kind of situation; the dispositions that turn given kinds of events into satisfiers or annoyers, into positive or negative reinforcers, for the subject or the species as a whole. The dispositions can be conceived as internal representations serving as standards against which actual situations are tested. These representations need not be explicit or reified or consciously accessible or consciously modifiable.” (Frijda 1986, p.335)

“Concerns are defined as internal representations of preferred states that serve as standards against which actual states of the world are tested. People seek to achieve them and events may agree or disagree with them.” (Frijda *et al.* 1991, p.213)

### **Embodied Artificial Intelligence**

New approach to studying Artificial Intelligence (AI) in the context of “complete” (embodied, situated) autonomous agents. It exploits the richness of behavior shown by an embodied agent that acts in the real world (as complex as it is) obtaining its (partial) information about the environment through its sensors in continuous interaction with the real world (situated agent). The development of Embodied AI has gone in parallel with “Behavior-Based Robotics”, the discipline that first pointed out the need to study intelligence in the framework of complete autonomous robots and that provides a natural test-bed for its theories.

### **Embodied Conversational Agent (ECA)**

“Embodied conversational agents are computer-generated cartoonlike characters that demonstrate many of the same properties as humans in face-to-face conversation, including the ability to produce and respond to verbal and nonverbal communication. They constitute a type of (a) multimodal interface where the modalities are those natural to human conversation: speech, facial displays, hand gestures, and body stance; (b) software agent, insofar as they represent their human users in a computational environment (as avatars, for example); and (c) dialogue systems where both verbal and nonverbal devices advance and regulate the dialogue between the user and the computer.” (Casell *et al.*, 2000, cover)

### **Emergence (emergent behavior, emergent functionality)**

“Emergence is a classical concept in system theory, where it denotes the principle that the global properties defining higher order systems or ‘wholes’ (e.g. boundaries, organization, control, ...) can in general not be reduced to the properties of the lower order subsystems or ‘parts’. Such irreducible properties are called emergent.” (Heylighen 1989)

“Agents can become more complex in two ways. First, a designer (or more generally a designing agency) can identify a functionality that the agent needs to achieve, then investigate possible behaviors that could realize the functionality, and then introduce various mechanisms that sometimes give rise to the behavior. Second, existing behavior systems in interaction with each other and the environment can show side effects, in other words, *emergent behavior*. This behavior may sometimes yield new useful capabilities for the agent, in which case we talk about *emergent functionality*. In engineering, increased complexity through side effects is usually regarded as negative and avoided, particularly in computer programming. But it seems that in nature, this form of complexity buildup is preferred.” (Steels, 1994). This notion is highly exploited by the new approach to Artificial Intelligence (AI) characterized as “Embodied AI”, “Bottom-Up AI” or Behavior-Based Robotics.

However one has to be careful not to mistake emergence for the unexpected effects produced by a lack of understanding of the system:

“We are often told that certain wholes are ‘more than the sum of their parts.’ We hear this expressed with reverent words like ‘holistic’ and ‘gestalt,’ whose academic tones suggest that they refer to clear and definite ideas. But I suspect the actual function of such terms is to anesthetize a sense of ignorance. We say ‘gestalt’ when things combine to act in ways we can’t explain, ‘holistic’ when we are caught off guard by unexpected happenings and realize we understand less than we thought we did.” (Minsky, 1986, p. 27)

### **Emotional Contagion**

“The tendency to automatically mimic and synchronize facial expressions, vocalizations, postures and movements with those of another person and, consequently, to converge emotionally.” (Hatfield *et al.*, 1992, pages 153-154)

### **Ethology**

The study of animal behavior under natural conditions, i.e., the animal’s responses are interpreted within the context of its actual environmental situation. Its aim is to interpret behavioral acts and whole patterns of animal behavior in ways that emphasize their functions and evolutionary history. Tinbergen (1963) categorized four areas of study in ethology: function, causation, ontogeny and evolution of behavior.

### **Goals**

“One hallmark of an active goal is that the individual will persist on the task, striving to reach the desired goal, in spite of obstacles and interruptions.” (Bargh and Chartrand 1999, p. 472)

“Once activated, a goal operates in the same way whether activated by will or by the environment” (ibid., p. 470)

“Goals do not require an act of will to operate and guide information processing and behavior. They can be activated instead by external, environmental information and events. Once they are put into motion they operate just as if they had been consciously intended, even to the point of producing changes in mood and in self-efficacy beliefs depending on one’s degree of success or failure at reaching the goal. The goal does not know the source of its activation and behaves the same way regardless of where the command to do its thing came from (...). Note

that this argument applies to complex self-regulatory goals - such as those that serve achievement motives - as well as to simpler behavioral goals.” (ibid., p. 473)

“The process of goal pursuit does not stop with the behavioral attempt to attain the goal, however. Inevitably, the individual either achieves or does not achieve (in varying degrees) the pursued goal and tends to evaluate his or her performance following the attempt. Many researchers have demonstrated the consequences of success or failure at conscious goal pursuit for one's mood and beliefs of self-efficacy (...). ... Our approach suggests that there are such consequences of succeeding and failing, even at goals of which one was not aware of pursuing.” (ibid. p.472)

### **Neural Network**

A Neural Network is a network of nerve cells (neurons) in an organism. Artificial Neural Networks (ANN) is the discipline of computer sciences that models those biological neural networks to use its computational properties. (cf. Rolls and Treves, 1998; Arbib, 2003)

### **Neuromodulation**

Neuromodulation refers to the action on neurons of a large family of chemicals called neuromodulators, e.g., dopamine, serotonin and norepinephrine. Each neuromodulator activates specific receptors on the neural membrane, having specific effects on the functioning of the neuron. Since neurons in different parts of the brain may have different receptors within its membrane, the same neuromodulator can thus have distinct effects in different parts of the brain. The overall result is that a single neuromodulator can modulate the functioning of a neural network. (cf. Kravitz, 1988; Fellous, 1999, 2004)

### **Perception-Action Model (or Hypothesis)**

“The Perception-Action Hypothesis (a term from motor behavior) is grounded in the theoretical idea, adopted by many fields over time, that perception and action share a common code of representation in the brain.” (Preston *et al.*, 2002) This hypothesis is closely related to the principle of sensory-motor coordination in Embodied AI and Behavior-Based Robotics, that states that all (intelligent) behavior is to be conceived as sensory-motor coordination that serves to structure the sensory input (cf. Pfeifer and Schreier, 1999).

### **Regulation (of emotions)**

Emotion regulation refers to a broad constellation of processes that serve to either amplify, attenuate or maintain the strength of emotional reactions. Included among these processes are certain features of attention that regulate the extent to which an organism can be distracted from a potentially aversive stimulus and the capacity for self-generated imagery to replace emotions that are unwanted with more desirable imagery scripts. Emotion regulation can be both automatic and controlled. Automatic emotion regulation may result from the progressive automatization of processes that initially were voluntary and controlled and have evolved to get generated in the absence of recruiting associated regulatory processes. For this reason, it is often conceptually difficult to distinguish sharply between where an emotion ends and regulation begins. Even more problematic is the methodological challenge of operationalizing these different components in the stream of affective behavior. (Davidsson 1999, p.104)

## Rule-Based System

A rule-based system is a particular instance of symbolic AI. As the name suggests, a rule-based system uses a library of operators or rules (e.g., of the form *If CONDITION(S) then ACTION(S)*) specific to a particular problem domain. Hence, the term ‘expert system’ describes a kind of rule-based system where the rules have been supplied by a human expert. An example of this is *Prospector*, an expert system used to assist geologists in locating valuable mineral deposits such as oil, coal or precious metals.

## Standards

Standards are a major determinant of the psychological significance of an event. A “standard” is a criterion or rule established by experience, desires, or authority for the measure of quantity and extent or quality and value. Both people and situations can be differentiated in terms of associated standards. Personal standards are seen to play an important role for individual differences in motivation, self-regulation and self-evaluation. Standards can function either as reference points or as regulatory criteria (e.g., tendency to surpass the performance of another person).

Standards constitute different kinds of knowledge – general declarative knowledge (social category standards), episodic knowledge (e.g., autobiographical standards), and procedural knowledge (e.g., normative guides).

Social standards are established by past interpersonal experiences, knowledge of self and others, and current social contexts. Action that occurs in relation to social standards is social action. (cf. Higgins, 1990)

## Symbolic Artificial Intelligence

Symbolic AI is best defined with the help of the classical water jugs problem: We have one 3-liter jug, one 5-liter jug and an unlimited supply of water. The goal is to get exactly one liter of water into either jug. Either jug can be emptied or filled, or poured into the other. One approach to implementing a solution would be to define a set of rules that encapsulate the behavior of water levels in the jugs after each action has been carried out. Consequently, the task can be regarded as the manipulation of the rules until the goal is reached, perhaps by depth-first search. Therefore, symbolic AI, as illustrated by this example, considers intelligence as problem solving that can be characterized by a set of rules and a method for manipulating them in order to satisfy some goal. Importantly, a result of this approach to AI is that the solution can be “human interpretable” – the solution is the sequence of rules applied to an initial state that solves the problem.

## Uphill Analysis and Downhill Invention (Braitenberg’s Law of)

“It is pleasurable and easy to create little machines that do certain tricks. It is also quite easy to observe the full repertoire of behavior of these machines - even if it goes beyond what we had originally planned, as it often does. But it is much more difficult to start from the outside and try to guess internal structure just from the observation of behavior. It is actually impossible in theory to determine exactly what the hidden mechanism is without opening the box, since there are always many different mechanisms with identical behavior. Quite apart from this, analysis is more difficult than invention in the sense in which, generally, induction takes more

time to perform than deduction: in induction one has to search for the way, whereas in deduction one follows a straightforward path. A psychological consequence of this is the following: when we analyze a mechanism, we tend to overestimate its complexity. In the uphill process of analysis, a given degree of complexity offers more resistance to the workings of our mind than it would if we encountered it downhill, in the process of invention.” (Braitenberg, 1984, page 20)

### **User Model**

A formal representation of the main characteristics of a user that may affect his/her interaction with software products or, more in general, with technology.

User models can have *static or dynamic components*. The static component includes a description of the long-term characteristics which are not likely to vary during interaction (typically: gender, name, social status). The dynamic component includes a description of those characteristics which are less stable (typically, knowledge, in interaction with intelligent tutoring systems). As far as characteristics with some affective connotation are concerned, the following is a list of the main ones, in decreasing order of stability: personality traits, values, norms, goals, preferences, mood, beliefs, intentions, attitudes, emotions.

### **User Modeling**

The method (and the software component that performs it) to build an initial user model and to update it consistently during interaction. Building and updating may be performed in an implicit or an explicit way. In *implicit user modeling*, data are acquired by the system without directly requesting them to the user. In *explicit user modeling*, data are acquired by direct interviewing. In both cases, some form of reasoning on the acquired data has to be done, to infer the user modeling features. An example of implicit acquisition in affective user modeling would be the “recognition” of the user’s emotional state from biological signals. An example of explicit acquisition: filling-up of personality questionnaires.

A key problem in user modeling is to insure consistency after *updating*. To this aim, typical *methods* of artificial intelligence may be applied: if the model is represented in logical form, truth maintenance and non monotonic reasoning methods are applied; if uncertainty is represented in the model, bayesian updating is the widely recognized appropriate method to apply. Other methods (like fuzzy logic, neural networks or learning algorithms) are appropriate in more specific domains and cases.

## 4 Review of key achievements in the thematic area

This section provides an overview of some of the key developments in the computational modeling of emotions in cognition and action. It is not a systematic review of the current state of the art but rather an illustration of the kind of computational systems that have been developed regarding different aspects of emotions in cognition and action. Emotion synthesis includes both, systems that “have” emotions and systems that reason about emotions, and mention will be made here to both types of systems. The interested reader is referred to the following publications for representative papers and surveys in the areas listed below: (Picard, 1997; Trappl and Petta, 1997; Cañamero, 1998, 2001b; Paiva, 2000; Cassell *et al.*, 2000; Cañamero and Petta, 2001; Carberry *et al.*, 2002; de Rosis, 2002a, 202b; Trappl *et al.*, 2003; Egges *et al.*, 2004; Hudlicka and Cañamero, 2004; Fellous and Arbib, 2004).

### 4.1 Emotion-based architectures for adaptive autonomous agents

Much research in emotion synthesis has been devoted to the design and implementation of emotion-based architectures for agents that must act in and adapt to changes in their environment autonomously. As it can be seen in the definition of “autonomous (and adaptive) agent” in Section 3, agents can be adaptive in different ways, depending on the temporal scale of the changes: adaptation to short-term, smaller changes in the environment involves flexible and rapid decisions, and is the problem dealt with by action selection architectures; adaptation to more significant and long-term (lasting) changes in the environment involves the ability to change and improve behavior over time and therefore learning. Let us briefly review some key achievements in these topics, illustrating them by a few representative examples.

#### Emotion in action selection

Computational models of emotions in this area have followed two main approaches: emotions “modeled” as emergent phenomena, or designed as an integral part of the agent architecture.

The “emergent” approach is typically adopted in artificial life computational models. As representative work in this tradition we can cite the “thought” experiments of Valentino Braitenberg (1984). His “Vehicles” – later implemented in real robots in numerous occasions – are simple machines or “robots” with very simple architectures consisting of direct connections between sensors and motors. They are situated in an environment containing a source of energy (e.g., heat, light) that can be detected by the sensors. By varying the connections between sensors and motors to be either lateral or counter-lateral, and the way the motors are powered by the sensors (either positively or negatively) different variants of a basic architecture can be formed that give rise to different behaviors; by varying the position from which the Vehicles approach the energy source, the same architecture will produce different behaviors and the same behavior can be observed in Vehicles with different architectures. Although these architectures do not implement any “emotional components” and the behavior of the Vehicles depends only on their morphology and interaction with the environment, the behavior displayed by some of these Vehicles can appear to an external observer as arising from internal states, and could be termed as “love”, “fear”, “aggression”, “cowardice”, or “curiosity.” In the same vein, Pfeifer proposed an artificial life environment of “Fungus Eaters” (1993) that also show emergent emotional behavior. This approach thus warns us of the risks of over-attribution and stresses the need to adopt a synthetic approach – by building

systems – in order to understand the mechanisms underlying (emotional) behavior. However, it also presents what can be seen as a major drawback (see Cañamero, 2003, for a discussion); modeling and implementing emotions as purely emergent phenomena in the eye of the beholder can be an exciting challenge for the designer of a robot, but this view misses a potentially important contribution of artificial emotional systems: their ability to relate to and influence different behavioral and cognitive subsystems at the same time.

The other approach postulates that, if emotions are to be meaningful to the robot, they must be an integral part of its architecture. However, to be meaningful, emotions must be grounded in an internal value system that is adaptive for the agent's (physical and social) environment, since it is this internal value system that is at the heart of the creature's autonomy and produces the valenced reactions that characterize emotions. As an example, the architecture proposed in (Cañamero, 1997) relies on both motivations and emotions to perform behavior selection. Initially implemented in simulated robots, this architecture is now being adapted to real robots (Avila-García and Cañamero, 2004). The robots inhabit a typical action selection environment containing various types of resources, obstacles that hamper their activities, and predators, and they must choose among and perform different activities in order to maintain their well-being (the stability of their internal milieu) and survive (remain viable in their environment, following (Ashby 1952) as long as possible — their ultimate goal. The architecture of the robots is behavior-based, and consists of: a *synthetic physiology* of survival-related variables controlled homeostatically (e.g., blood sugar, vascular volume, energy, etc) and “hormones” that can alter the levels of the controlled variables; a set of *motivations* (aggression, cold, curiosity, fatigue, hunger, self-protection, thirst, and warm) activated by “errors” (deficit or excess) in the levels of the controlled variables when these depart from their ideal values, therefore setting the internal needs of the robot; a repertoire of *behaviors* that can satisfy those internal needs or motivations (and also create new ones), as their execution carries a modification (increase or decrease) in the levels of specific variables; and a set of “basic” *emotions* (anger, boredom, fear, happiness, interest, and sadness) that can be activated as a results of the interactions of the robot with the world — the presence of external objects or the occurrence of internal events caused by these interactions — and release “hormones” when active. Under “normal” circumstances, behavior selection is driven by the motivational state of the robot. Emotions constitute a “second order” control mechanism running in parallel with the motivational control system to continuously “monitor” the external and internal environment for significant events. They can alter motivational priorities and behavior execution through the effect of released hormones on the physiology, arousal, attention, and (internal and external) perception of the robot. Closely related architectures have been implemented in robots by Velásquez (1998) in the case of action selection and learning tasks, and by Breazeal (2002) in a social robot.

The architectures described above incorporate very simple and low-level cognitive mechanisms. An example of a much more complex architecture, also integrating high-level aspects of cognition influencing lower ones is the three-layered framework developed by Aaron Sloman (2003). This architecture was not developed only for the purpose of action selection but is much more general. It includes three layers with very complex interactions and ‘loops’: a reactive layer, a deliberative layer, and a meta-management layer. Paralleling this organization, Sloman distinguishes between primary, secondary, and tertiary emotions.

## Learning

Learning in autonomous agents (in particular in robots) typically follows association or reinforcement models, and makes use of some kind of external signals (arising from the

environment or supplied by a “critic”) that provide positive or negative reward. Computational models of emotion-related learning in autonomous agents have also generally adopted this type of approach. Examples of systems that use reinforcement learning are provided by the work of Gadanho, that uses a neural network architecture and reinforcement learning to implement an adaptive robot controller that learns to navigate in environments of varying level of difficulty (Gadanho and Hallam, 2001). The MITB architecture of Ventura (Ventura et al., 2001), closely inspired by the “movie-in-the-brain” idea introduced by Damasio (1999), uses reinforcement learning to select courses of action aiming at obtaining desirable states for an agent and learn to perform a pendulum balancing task.

One of the main problems underlying traditional reinforcement learning models is how to make those signals truly meaningful to the robot so that the learning process is more autonomous and grounded in the architecture of the robot. In other words, how can a robot make sense of the perceived signals by itself, as opposed to using reward information provided by some sort of “external teacher”? How can it decide what to learn and what *not* to learn? A simple model of learning addressing this issue was proposed by Blumberg (1996), who developed an animated dog, Silas T. Dog, modeled closely after ethological models of animal behavior; this character possesses internal variables that represent emotions (as well as motivational states such as hunger or thirst) and that can influence what he learns. Changes in the dog’s internal variables drive a learning process, e.g., when the ‘fear’ variable increases, Silas tries to determine which stimuli from his perception and short-term memory can best predict the change, and use that association to learn new ways to behave, e.g., avoid places where he previously perceived objects that caused ‘fear’. In this work, however, emotions are modeled in a very simplistic way as simple variables. A more complex approach is exemplified by the work of Velásquez, who implemented fear conditioning in his robot Yuppy (Velásquez, 1998b) using an associative network model. The architecture of Yuppy contains, in addition to perceptual, behavior, motor, and drive systems, a set of basic emotions. Simulated “pain” signals, triggered when a person disciplines Yuppy, allow it to learn “secondary” emotions, and in particular to associate a new cognitive releaser (the sound of a flute) when this is paired with the releaser that caused “pain”.

A more biologically plausible approach would need to include a mechanism rooted in an internal “value system” to provide internal signals regarding the “positive” or “negative” qualities of actions and stimuli, giving them a meaning with respect to the values, needs and goals of the robot beyond a metaphoric use of the terms “pain” and “pleasure” to refer to punishment and positive reward, and allowing to learn appropriate valenced reactions to them. The work of (Andry *et al.* 2001) and (Cos-Aguilera *et al.* 2003) provide initial solutions in this direction, where “pain” and “pleasure” signals are rooted in discomfort or well-being related to the stability (or lack of it) of other internal variables of the robot, and are used to learn affordances as a result of interactions with the environment.

## Memory

Finally, management of memory is another major problem in autonomous agents (in particular robots) that have to timely select appropriate courses of action. If the robot lacks appropriate criteria to filter out information, its memory is then too global, causing problems of cognitive overload and very long recall times. Mechanisms for selective memory inspired from emotional memory in humans (e.g., phenomena like mood-congruent recall of past memories) and the related notion of autobiographic memory, can help to solve some of these problems and also provide generally coherent responses to a wide range of situations. The work of Araujo (1994) addressed this problem. His model, inspired by the theories of LeDoux about

the dual route of emotion processing (LeDoux, 1989, 1996), tries to integrate low-level physiological emotional responses and their high-level influences on cognition by using two neural networks – “emotional” and “cognitive”. Interactions between these two networks imitate mood-congruent memory retrieval and learning, and the effects of anxiety on memory performance.

## 4.2 Appraisal systems

In the computational domain, appraisal theories have given important inputs to both theory modeling and emotion synthesis and, in the last two decades of the 20<sup>th</sup> century, they provided a decisive boost to the field. At the same time, theoreticians have also been raising critical voices of importance to keep engineering efforts on track and within the limits of realistic achievability (e.g. Kaiser and Wehrle, 1994; Wehrle, 2001). The early 1980s witnessed a marked increase of interest in the computational area, with research in natural language systems progressing from issues of text understanding and generation and knowledge representation of a more basic nature to more application-oriented domains of story-telling, story-understanding, and dialog systems and the rise of situated interaction as research topic (in both, virtual and robotic settings; see e.g. (Frijda, 1995) for an early repercussion on theory). Achievements and insights from the community with descent from natural language systems research are reflected in influential models such as the theoretical OCC model (Ortony et al., 1988) and numerous implementations, beginning with the Affective Reasoner architecture (Elliott, 1992) that builds upon and extends the representation developed for the BORIS system (1981) and was applied to a number of application domains, including educational settings and embodied conversational characters (Elliott 1997, 1997b) and the Em component (Reilly & Bates 1993, Reilly 1996) of the Hap architecture (Bates et al., 1992; Loyall, 1996), which has evolved further into a component of commercial products (e.g. Zoesis Inc.). With the increase in variety and number of the deployment scenarios and its frequent application to emotion synthesis tasks for which it was not originally intended (e.g. Martinho and Paiva, 1999), the OCC model itself has been streamlined and extended towards fuller coverage of the first-person perspective, including different dimensions of coping and explicating the connection to personality models (Ortony, 2003); at the same time, research on methodological requirements and engineering aspects of OCC-based implementations continues (e.g., Gebhard et al., 2003).

Among other bodies of work in appraisal theory, Frijda’s model has been adopted (and adapted) in a number of systems, including ACRES (Frijda & Swagerman, 1987), WILL (Frijda and Moffat, 1994; Moffat 1997), and TABASCO (Staller and Petta, 1998; Petta, 1999; Staller and Petta, 2001; Petta, 2003). The latter effort is characterized by an attempt to analyze and integrate the insights and salient points from various theoreticians, including (Leventhal and Scherer, 1987; Smith and Lazarus, 1993; Roseman *et al.*, 1996; Keltner and Gross, 1999; Reisenzein, 2001) in a process model that also captures insights from software engineering and artificial intelligence research. In a similar vein, Gratch (1999, 2000) re-interprets appraisal theory from an engineering perspective, identifying interesting complementarities between generation and perception of planful action; for recent evolutions of this model, see e.g. (Marsella and Gratch, 2003). Steve Allen carried out another noteworthy effort of re-synthesis, developing a comprehensive control architecture for autonomous agents centered around the central motivating construct of concerns (Allen, 1999). Velásquez (1996) implemented the cognitive reasoning component of the Cathexis model via an adaptation of the theory covered in (Roseman *et al.*, 1996). Scherer (1993) presents an attempt to formalize his sequential stimulus evaluation check theory in the form of an expert system subjected to empirical evaluation, a line of research furthered in e.g. (Wehrle and Scherer, 2001; Kaiser

and Wehrle, 2001). Finally, Hudlicka's appraisal-based architecture, MAMID, has modeled other aspects such as affect regulation and induction of emotions in others (Hudlicka, 2001), and the interacting influences of states and traits on perception and cognition, including their effects on the appraisal process itself (Hudlicka, 2004).

The OCC model to date clearly remains a key reference for the development of applications - in particular in the domain of embodied conversational characters. At the same time, experiences gained with the attempts to utilize the model in a broad range of application domains is helping to understand the delineations of the actual scope of the model in terms of suitable application domains and functional coverage, defining directions for further research. One important such direction is to overcome the significant simplifications implemented in the large majority of systems deployed today, which adhere to a strictly linear sequencing in processing with little of the explicit and contextual bi-directionalities and reciprocal influences emphasized in the theory. To overcome this impasse, exploiting modern resources from engineering theory, modeling approaches, is one prominent candidate topic for collaboration between WP3 and WP7 of straightforward relevance for all other application-oriented workpackages. Another important move would be to further research into models of fluid emotions, as already attempted in some of the implemented process models mentioned, to complement the predominant exclusive consideration of crystallized emotions.

In spite of these limitations, and following the eminent example of Clark Elliott (arguably the first leading figure in terms of understanding, discussing, and proposing required extensions to the original OCC model for different application scenarios), the results gathered to date in efforts to translate theoretical insights from different appraisal theories into components of practical value provide an important platform to both improve and build upon. A platform that - if clearly in need of critical inspection and assessment - is apt to serve as basis and connecting hub for research extending into a multitude of directions, from low-level issues such as constituents of perception to high-level issues of conceptual processing, from the individual to the social perspective.

### 4.3 User modeling

Although the influence of affective factors in human-machine interaction has been recognized since long – see the various studies by Nass and his group (Reeves and Nass, 1996) and (Picard, 1997) – methods and experiences aimed at including personality, emotions and attitudes among the characteristics which are recognized, interpreted and formalized in user models are more recent. Proceedings of the Workshops on “Attitude, personality and emotions in User-Adapted Interactions”, organized in the scope of the User Modeling Conferences '99, '01 and '03 witness an increasing interest in this area, as well as the following Special Issue of the journal *User Modeling and User-Adapted Interaction* (de Rosis, 2002).

In the majority of applications (like intelligent tutoring systems or museum tour guide), “attitudes” like frustration, anxiety, doubt, cognitive load, fatigue and level of attention of the user are important to consider, as well as their positive counterparts (satisfaction, enthusiasm, and so on). In other cases (typically, affective dialogs), stronger emotions may occur, which the system should be able to recognize, interpret and formalize in the user model, consistently with other mental components.

Integration of “recognition” methods with their “interpretation” is essential: an attitude or an emotional state may be due to several causes, and guessing the most likely of them is a starting point, for the system, to react appropriately. To this aim, the system should try to

figure out which was the combination of mental components which might have produced the user state: beliefs about goal achievement or threatening, respect of values and norms and so on. Methods employed in the two phases of recognition and interpretation are not the same: essentially signal or image analysis in the first step, versus some form of logical or uncertain reasoning (typical of AI methods) in the second one (see, e.g., Ball, 2003). This shows the need and offers the opportunity to integrate work done in WP4 (about recognition methods) with activity of WP7 in the area of cognitive models.

Once the user affective state has been recognized and interpreted, the system must decide “how to react” to this situation: in other words, the level, type and form of empathy it should show in the considered situation. Knowing the reasons why (probably) the user “feels” a given emotion will help it in reacting appropriately. If employed in a “simulative” way, the user model will enable the system to find out appropriate strategies to get the desired behavior in the user: in particular, to induce an emotional state which may be instrumental in achieving a system goal. For symbolic approaches, the ground-breaking work on the Affective Reasoner (Elliott, 1997; Elliott et al., 1999) remains to date a major reference. Current research trends aiming to exploit richer and dynamical contextual (e.g., including causal and historical) knowledge available to the system (as in the efforts by Jonathan Gratch et al. at ISI). Likewise, subsymbolic efforts start to include components such as policy boxes (Ball, 2003).

Formalization of empathic reaction strategies offers, again, the opportunity of integrating work done in this WP with methods and exemplars developed in the scope of WP8, about affective dialogs, persuasion, humor and politeness: these may be seen as ways of “showing an empathic behavior” which requires, to be effective, a clear image of the user attitude and mood, and therefore a consistent and dynamically updated user model.

#### **4.4 Embodied conversational agents and virtual environments**

Embodied Conversational Agents (ECAs) and Virtual Environments (VEs) provide a means to simulate existing models for emotional action visually. With the emergence of 3D graphics, we are now able to create very believable 3D characters that can move and talk. Multi-modal interaction with such characters is possible as the required technologies are getting mature (speech recognition, natural language dialogues, speech synthesis, animation, and so on). Within HUMAINE, ECAs and VEs offer a flexible platform and testbed to validate whether or not a model for emotional cognition/action is giving desirable results. Research so far has been undertaken in the following topics:

##### **Emotions, personality, and behavior**

The effect of emotions and personality on behavior and action has been extensively researched (Piwek, 2002), whether it concerns a general influence on behavior (Marsella and Gratch, 2002), or a more traditional planning-based method (Johns and Silverman, 2001). Various rule-based models (André *et al.*, 2000), probabilistic models (Ball and Breese, 2000; Kshirsagar and Magnenat-Thalmann, 2002) and fuzzy logic systems to model the triggering of emotions due to events (El-Nasr *et al.*, 2003) or to map facial expressions of an emotion with a given intensity (Dui Buy et al, 2003) have been developed. In (Ball and Breese, 2000) the authors developed a model in which the emotion an agent is undergoing may affect her verbal and non-verbal behavior. They built a Belief Network that links emotion with verbal and non-verbal manifestation. de Rosis and colleagues have developed a computational model of emotion triggering using a dynamic Belief Network (Carofiglio *et al.*, 2004). Their model is

able to determinate not only which emotion is triggered after a certain event for a given agent but also it is able to compute the variation of this emotion over time: this emotion may increase or decrease in intensity or it may also evolve in another emotion. The computational model uses a Belief Desire Intention (BDI) model of the agent's mental state. Carmen's Bright IDEAS (Marsella et al., 2000) is an interactive drama where characters exhibit gestures based on their emotional states and personality traits. Through a feedback mechanism a gesture made by of a character may modulate her affective state. A model of coping behaviors has been developed by (Marsella and Gratch, 2003). The authors propose a model that embeds information such as the personality of the agent and his social role. In these latter works, behaviors are viewed as responses to events and actions: what is simulated is how the agent responds with an emotion to what happens in his own mental state, and how an emotion affects the agent's behavior.

### **Communicative Behaviors**

In the construction of embodied agents capable of expressive and communicative behaviors, an important step is to reproduce affective and conversational facial expressions on synthetic faces. Cosmo (Lester et al., 2000) is a pedagogical agent particularly keen on space deixis and on emotional behavior: a mapping between pedagogical speech acts and emotional behavior is created by applying Elliott's (1992) theory. André et al. (2000b) developed a rule-based system implementing dialogs between lifelike characters with different personality traits (extroversion and agreeableness). Cassell and her colleagues have proposed several ECAs' settings. REA, the real estate agent (Cassell et al., 1999), is an interactive agent able to converse with a user in real-time. REA exhibits refined interactional behaviors such as gestures for feedback or turn-taking functions. Cassell and Stone (1999) designed a multi-modal manager whose role is to supervise the distribution of behaviors across the several channels (verbal, head, hand, face, body and gaze). Recently, Cassell et al. (2001) built a flexible and extensible toolkit, BEAT, to synchronize verbal and nonverbal behaviors. To ensure higher portability of the system, XML is used as a tagging scheme for any input coming from the dialog manager and for any output to be sent to the animation module. BEAT tags the input text with linguistic and contextual information from which behavior computation is done. This computation is obtained by specifying rules relating meaning to signals. BEAT allows one to add any rules easily. Pelachaud et al. (2002) have created an ECA, Greta, that incorporates communicative conversational aspects. To determine speech-accompanying non-verbal behaviors the system relies on a taxonomy of communicative functions proposed by Poggi (2001). The representation language 'Affective Presentation Markup Language' (APML) is used to control the agents behavior. The system takes as input the text (tagged with APML) the agent has to say. The system instantiates the communicative functions into the appropriate signals. The output of the system is the audio and the animation files that drive the facial model.

### **Social context modeling**

The role of social context in an agents behavior has also been considered. DeCarolus et al., (2001) propose a model that decides whether an agent will display or not her emotion depending on several contextual and personality factors. Prendinger et al. (2002) integrate contextual variables, such as social distance, social power and thread, in their computation of the verbal and non-verbal behavior of an agent. They propose a statistical model to compute the intensity of each behavior. Rost and Schmitt (2002) modeled how social relationship and attitudes toward others affect the dynamism of an interaction between several agents.

## 5 Review of key problems in the thematic area

In humans, emotions entail distinctive integrated ways of perceiving and assessing situations, processing information, and modulating and prioritizing actions. Elaborating computational models that embed the effects of emotions in cognition and action is a complex, multi-faceted problem that poses multiple integration challenges at both conceptual and computational / technical levels. This section reviews some of these problems.

### 5.1 Conceptual problems

Section 2.2 (Key issues) enumerated some of the key topics that should ideally be addressed within this Workpackage. This section develops problems underlying some of those topics that we believe should be addressed in priority within the exemplars.

#### 5.1.1 Mechanisms underlying the involvement of emotions in cognition and action

Within the area “Emotion in Cognition and Action”, a key issue that must obviously be addressed is an investigation of possible mechanisms that allow emotional phenomena to influence cognitive and behavioral processes and how they can be implemented in artifacts. Different emotion theories and models (see the theoretical models of emotions listed under WP3 of the deliverable D2a and in Section 4 of D3c) put the emphasis on diverse aspects of emotional phenomena and the different mechanisms underlying them, and this often implies different ways to conceptualize the link between emotion and cognition/action. For example, some models such as “circuit models” (e.g., Panksepp, 1998; Rolls, 1999) and “adaptational models” (e.g., LeDoux, 1996) postulate specific “emotion centers” or “neural circuits” in the brain, particularly or primarily concerned with the processing of emotion-related information and the production of emotional responses. Other models, related to “peripheral feedback models” (e.g., Damasio, 1999), such as (Fellous, 1999, 2004) consider emotions as dynamical patterns of neuromodulations, rather than patterns or circuits of neural activity. These approaches clearly entail different ways of conceptualizing the link between emotions and cognition and action. In the former case, it involves establishing connections between specialized “emotion circuits” and other brain circuits and areas primarily involved with different cognitive and behavioral functions. In the latter, patterns of neuromodulations directly affect brain areas involved at all levels of (cognitive and behavioral) functions, and the extent to which emotions can affect different aspects of cognition and action depends on the potential for neuromodulation of the neural substrate involved in those different aspects. The operationalization of these approaches will give raise to different computational models and mechanisms. For example, a computational model inspired by “circuit” and “adaptational” models such as (Balkenius and Morén, 2001; Morén, 2002) typically uses neural networks to implement different brain areas in the “emotion circuit” such as the amygdala, thalamus, sensory cortex, and orbitofrontal cortex, and their interconnections in a “cognitive task” such as learning (emotional conditioning in this case). On the contrary, a computational model inspired by a “neuromodulation” model such as (Cañamero, 1997; Avila-García and Cañamero, 2004) would not model explicitly “emotion centers” but rather models different emotions in terms of simulated “hormones” that affect the functioning of different elements of the architecture, modulating different cognitive and behavioral functions

such as perception, attention, motivational priorities, behavior selection, and behavior execution.

### 5.1.2 Emotion elicitors

The study of mechanisms underlying the involvement of emotions in cognition and action in computational models does not necessarily shed light regarding what mechanisms must be in place for “external” and “internal” influences to activate or produce an emotion in the first place. Researchers in the computational modeling tradition often cite Izard’s theory (Izard, 1993) of four emotion elicitors – neural / neurochemical, sensorimotor, motivational and cognitive – but very few have attempted to implement the four elicitors simultaneously. Some of the problems that their implementation involves are, for example, establishing the causal relations among the different elicitors, or deciding which computational approaches and techniques are better suited to implement each of them and how they can be integrated. Appraisal theories are another approach that has become a very popular source of inspiration for computational models, in particular the more “cognitive-oriented” ones. However, the two problems mentioned above reappear in most cases since, with few exceptions (e.g., Scherer, 2001; Smith, 2004) the gap between the level of abstraction of the theory and the concrete decisions needed for their implementation is too big, and engineers are left to their own intuitions in the search for solutions to bridge that gap – solutions that, understandably, are often ‘ad hoc’ and driven by the particular needs of the application. Guidelines are needed for the operationalization of these theoretical models, and their elaboration necessitates a joint collaborative effort between theorists and computer scientists.

### 5.1.3 Emotions as cognitive modes

In humans, emotions entail distinctive integrated ways of perceiving and assessing situations, processing information, and modulating and prioritizing actions. In this respect, emotions can be seen as different “cognitive modes” that have a “global” and synchronized influence in our perceptual, cognitive, bodily and behavioral relation with the world. Achieving this in computational emotion architectures involves a number of challenging problems, many of them largely unexplored, such as:

- Which aspects of cognition need to be in place in the architecture to be able to speak of a “cognitive mode”?
- What mechanisms can be used to modulate different aspects of perception and cognition?
- What mechanisms are required to implement the different effects of various emotions?
- How can computational models take into account cultural and individual differences in the synthesis of emotions as “cognitive modes”?
- Which are the causal relations between the different subsystems involved?
- Which (computational) mechanisms allow the integration of the “fast” and “slow” pathways in the processing of emotion-relevant information?

- How to synchronize the effects of emotions on the different computational (cognitive, bodily and behavioral) subsystems involved in order to obtain the “global” effect that emotions have in humans?
- How can we model the influence that emotions have in the perception of the social partners (e.g., to take this into account in a user model)?
- In biological systems, “one of the main functions of emotion is to achieve a multi-level communication of simplified but high-impact information” (Fellous, 2004, p. 39). However, on what grounds can an autonomous artifact assess what constitutes “high-impact” information? In other words, what kind of mechanisms (value systems) are needed to make information “emotionally relevant” for an autonomous artifact? How can “value systems” be implemented as an integral part of the architecture in order to ground “emotional meaning” for the artifact?
- How can we model the relation between the “cognitive modes” and the action tendencies involved by emotions in different architectures?

#### **5.1.4 Emotions, value systems, motivation, and action**

These notions are closely related in multiple ways, and different computational models have implemented these different aspects. In emotion synthesis, emotions play important roles in relation to the production of action in autonomous artifacts (see e.g., Cañamero, 2003), e.g.:

- The fact that emotions are related with (“general”) goals or concerns, rather than with particular behavioral response patterns (emotions versus goal-directed behavior), explains the fact that they allow to generate richer, more varied and flexible behavior.
- They constitute “second-order” control mechanisms that constantly monitor the internal and external environment to detect and respond to potential “threats” of different sorts, therefore either sustaining or interrupting ongoing goal-directed behavior.
- They modify/amplify motivation, producing changes in motivational / goal priorities to deal more efficiently with certain types of relevant (survival-related) events.
- They can also constitute motivational factors and constitute “value systems” that affect the selection of goals and goal-directed behavior.

The implementation of these functions in computational architectures poses, once more, non-trivial problems regarding the choice of underlying mechanisms and the integration of motivational, behavioral, and emotional components of the architecture. In addition, the link among these components must be grounded in some sort of “value systems” to permit the autonomous generation of valenced reactions that characterize emotions and distinguish them from “cognitions”. A no minor problem is that of assessing (measuring quantitatively) the benefits that the effects of emotions in the generation/selection of behavior carry for the performance of the artifact. This needs the development of different performance indicators and a systematic understanding of the problems implicit in different types of environments.

## 5.2 Integration Challenges

The task of WP 7 presents considerable integration challenges at various levels:

### 5.2.1 Problems arising from theories and models of emotions in humans and animals

Sound research into computational models of emotions needs inspiration from theoretical emotion models; however, researchers in the computational tradition often get bogged down when confronted with the vast literature in the psychology, neuroscience and sociology of emotions. As put forward in deliverable D3c of WP 3, existing theoretical models of emotions are very diverse, put the emphasis on very different aspects of emotions, often have different scope, and do not even seem to agree on a common definition of the phenomena they are talking about. The same problem re-occurs in existing computational models, which in addition need a more sound understanding of their theoretical foundations and of how theoretical models can be meaningfully mapped into computational mechanisms underlying emotional behavior. Much conceptual and theoretical ambiguity and confusion needs thus to be resolved, and in fact collaborative work between theorists (WP3) and computational modeling researchers (WP7) might contribute to such clarification task, since notions have to be precisely defined to be operationalized and implemented.

A no minor problem that also requires pluridisciplinary collaboration stems from the fact that researchers in computational modeling only have a partial understanding of theories in psychology, neuroscience, etc. Even when a good knowledge of the literature is achieved, without the guidance of researchers that have solid hands-on experience in these areas it is difficult to achieve a deep understanding of a theory and, in particular, to discern how consolidated different notions and results are, what their real complexities and implications are, to what extent they can be “believed”, and what kind of support they can provide to computational modeling.

Finally, a big gap exists between theories and implementations regarding which aspects of emotions are relevant to capture in the theory or model; what seems essential to the theorist might sound like a vacuous concept to the computer scientist, and what the computer scientist captures in operational models might seem unacceptable trivial to the theorist. Only a collaborative effort can provide sound answers to the question of how can theoretical models be meaningfully mapped into computational mechanisms underlying emotional behavior.

### 5.2.2 Problems with computational models, representation formalisms and implemented systems

The question of how can theoretical models be meaningfully mapped into computational mechanisms underlying emotional behavior does not have an easy solution or a single answer. Diverse computational approaches provide different conceptual frameworks, techniques, representations and systems to achieve these mappings, posing additional integration challenges at all these levels.

Computational approaches relevant to model emotions are offered by disciplines as diverse as Symbolic Artificial Intelligence, Embodied Artificial Intelligence, Behavior-Based Robotics, Autonomous and Synthetic Agents, Cybernetics, Virtual Environments, Artificial Life,

Dynamical Systems, etc. In some cases, these disciplines cover different but complementary aspects; however, in many other cases they regard themselves as alternative and conflicting paradigms. A clear example might be provided by the classical debate between Symbolic and Embodied Artificial Intelligence, which adopt very different perspectives regarding their definition of intelligence and the problems that are relevant for its investigation. Symbolic AI traditionally equates intelligence to reasoning capabilities and, adopting a dualistic stance in the old philosophical mind-body problem, postulates that intelligence must be studied independently from any sort of embodiment. Intelligence can be divided in different “modules” or cognitive functions (e.g., planning, learning, memory, etc.) that can be studied independently. Symbol systems are the best representation since they provide a unified, general formalism underlying all the different aspects of intelligence; their meaning is attributed by the human designer of the system and computation consists in symbol manipulation following logical rules drawn from introspective reasoning. When it comes to emotion modeling, Symbolic AI is primarily interested in the involvement of emotions in reasoning and “high-level” cognitive processes such as attribution of beliefs, desires and intentions, reasoning about emotions and cognitive appraisal of emotionally significant events, objects and situations, or logic-based decision making and planning. Contrary to this, for Embodied AI intelligence cannot be understood independently from its embodiment and the (physical and social) environment in which it is embedded; intelligence is not only nor primarily reasoning, and there is no “golden standard” for intelligence, but many different types of intelligence (with different degrees and types of complexity) are equally valid as they are relevant for different types of environments and interactions. Intelligence can thus only be studied in “complete creatures” in interaction with their environment; the emphasis is not on the study of independent cognitive functions but rather in the adaptive value that different morphologies, architectures and mechanisms provide to survive in particular environments. Appropriate underlying representations are multiple (and usually non-symbolic ones), as far as they allow the “complete creature” to build its own “view” of the world as a result of the dynamics of its interactions with it. With respect to emotion modeling, Embodied AI is primarily concerned with the embodiment aspects of emotions, their adaptive value, and their (modulatory) effects on “low-level” cognition and action. Although the philosophical views underlying Symbolic and Embodied AI are very different and, to a big extent, incompatible, the fact that they are addressing very different aspects of intelligence and emotion makes their actual scope and practical achievements somewhat complementary. Taking a pragmatic approach that tries to reconcile “the best of both worlds”, some researchers have attempted to integrate these two paradigms in so-called “hybrid” architectures. Such architectures typically combine a “reactive” module inspired from Embodied AI systems to carry “low-level” cognitive and behavioral tasks and a “symbolic” module to perform reasoning and “high-level” cognitive tasks, whereas integration is usually handled by an intermediate layer that gives priority to the “lower” or the “upper” layer and in some cases performs some sort of integration, although this is in many cases done in a rather ‘ad hoc’ manner. In the case of emotion, a good example of a hybrid architecture is the “Cogaff” 3-layered framework developed by Aaron Sloman over the last decades (Sloman, 2003). Understanding of the scope, adequacy, and potential for cooperation/integration of different conceptual traditions and modeling paradigms is therefore the first step in any integrative effort towards the elaboration of sound and comprehensive emotion-oriented systems.

A sound theoretical understanding of these conceptual traditions does not provide a complete solution to the integration problem, however. Representation formalisms and techniques stemming from those traditions are equally varied: rule-based systems, frames, scripts, neural networks, finite state machines, artificial chemistries, artificial immune systems, artificial hormonal systems, mathematical models, mark-up languages, etc. The different formalisms

and techniques are more or less appropriate to model particular aspects of emotions. An understanding of their scope, adequacy, and potential for cooperation / integration is therefore a necessary first step. An emotion-based architecture will normally include different aspects of emotions and different cognitive and behavioral functions, therefore requiring the integration of the various formalisms and techniques underlying those aspects and functions. In the simplest cases, this integration might be achieved via the input / output data flow among the various modules of the architecture, following a “black box” modeling approach (see Wehrle and Scherer, 1995; Wehrle, 2001 for a characterization and discussion of “black-box” versus “process” modeling approaches). In many other cases, however, a deeper integration will be necessary, in particular if we adopt a “process modeling” approach that attempts to simulate naturally occurring processes, since in this case the causal factors affecting the underlying mechanisms involved and the structural dependencies among these mechanisms need to be modeled explicitly. This modeling approach might thus require the deep integration of very different formalisms. For example, a neuromodulatory emotion model such as that mentioned in Section 5.1.1 might be suitably implemented using simulated hormones and artificial chemistries; in the same architecture, perceptual processes might be implemented using a neural network and a planning or decision making module might be implemented by a rule-based system. Emotional modulation of perception and decision making would imply the far from trivial tasks of integrating a simulated chemistry simultaneously with a neural network (for perception) and with symbolic rules (for decision making).

The above problems re-occur with existing systems. Implemented systems often address partial and complementary aspects of emotional phenomena; therefore, their integration would allow to come up with much more complete and self-contained systems. This integration entails tackling all the conceptual and technical integration problems mentioned above. A different but related challenge is posed in the case of systems that implement overlapping (rather than complementary) aspects, or even provide different implementations of the same underlying type of theoretical model (e.g., various systems implementing different appraisal models). Much understanding could be gained from accurate comparisons and evaluations of such tools; for this, reusable comparison and evaluation criteria, methods, tasks and scenarios need to be developed. Within HUMAINE, this is a route that should ideally be explored in collaboration not only with WP 3 but also with WP9 (Usability) and WP10 (standards).

## 6 Assessment of the key development goals in the thematic area

### 6.1 Some current “bottlenecks” that need to be addressed

As illustrated in Section 4, numerous computational emotion-based models and systems have been developed in recent years. In spite of the thriving growth of this area and the efforts by many to develop solid pieces of work grounded in sound research, a number of problems constitute, at present, “bottlenecks” that can only be addressed by large-scale, long-term pluridisciplinary efforts. Many questions need to be answered in order to achieve principled emotion-based architectures that at the same time provide (a) meaningful solutions to problems arising from interactive systems and autonomous agents research, and (b) useful tools for emotion theorists; some of these questions are:

- *Regarding models:* What is the scope of the different types of emotion theories and models? Do they explain the same phenomena / aspects of emotions? To what extent (and which ones) are they incompatible / can they be combined? What is the notion (definition) of “emotion” underlying each of them, and what sort of consequences and “constraints” does each definition put regarding the operationalization and implementation of each model? Is a general definition of emotions possible / required for modeling?
- *Regarding emotion “machinery”:* Which are plausible mechanisms underlying different aspects of emotions and their influence in cognition and action? What kinds of conceptual and computational mechanisms are better suited to explain and model the relations between emotion and cognition-action? How can computational mechanisms stemming from different conceptual traditions be integrated?
- *Regarding applications:* Which emotions / aspects of them can be meaningfully implemented in artificial autonomous creatures? What guidance can emotion theories and models provide in the search for answers to this question? (e.g., focus on the adaptive value / functions / components of emotions?) How can different models be suitably operationalized / for which applications?
- *Regarding the assessment of the influences of emotion in cognition and action:* How can emotional states/processes be quantified and measured (a) from “inside” the organism / agent architecture; (b) from observed behavior? To what extent does “emotional behavior” respond to the effects of specific internal “emotion machinery”? To what extent can it be explained as a “side effect” of the interactions of an agent with its environment? (emotion as a notion in the eye of the beholder). To what extent does the analysis of observed behavior help us understand underlying emotional mechanisms?

We don’t think that these questions need to, or indeed can be (fully) answered from a theoretical perspective prior to computational modeling and implementation. We rather see the interplay between theory and computational modeling as a two-way, multidisciplinary enterprise in which modeling can provide support and insights in the search for those answers, and in some cases even give rise to the reformulation of some theoretical questions and problems. Within the context of HUMAINE, attempts to tackle those problems necessitates collaborative work not only between the members of WP7 but across workpackages. It would

be unrealistic to think that solutions will be found during the life time of the project. However, HUMAINE can greatly contribute to the advancement of the field by bringing the need to solve these problems to light and by taking some initial steps to approach them. The exemplars will thus be oriented toward this objective.

## 6.2 Key development goals

Taking into account the need to dissolve the above mentioned bottlenecks, and extrapolating from the key developments and problems presented in previous sections, this section outlines some research directions that constitute key development goals in the area. This section also builds on the arguments presented in (Cañamero, 2001a, 2001c; Picard, 2003; Wehrle, 2001). Once more, this section does not intend to provide an exhaustive account of key development but rather focuses on topics particularly linked to the exemplars currently foreseen.

### 6.2.1 The “origins” and grounding problem of artificial emotions

To date, the design of most synthetic emotional systems is based on the intuitions of the designers of those systems regarding the choice of primitives or building blocks (emotion “components”, “modules” or “features”) that constitute the emotional system. In most cases those building blocks are “hardwired” (design manually) by the designer, taking inspiration from characteristics and functions of emotions in humans; such “building blocks” are also labeled after terms used in the psychology and neuroscience of human emotions. While this approach, based on elements and terms already familiar to us, fosters understandability of the system, it carries two major dangers: (1) the risk of “over-attribution” by human users of functions and capabilities implied by those terms that humans possess but artifacts do not have (a good example being the attribution of “feelings” and conscious subjective states to artifacts); (2) the lack of “grounding” of those emotion components, since they were included in the system on the grounds of their meaning for the human designer and user rather than due to their meaning for the artifact in interaction with its (physical and social) environment. Within this approach, avoiding the risk of over-attribution requires honesty and transparency by the designer regarding the mechanisms underlying these systems, in particular when presenting work to the end user and the layperson. Within HUMAINE, this is an issue that must be considered under WP10 (Ethics and good practice). As for emotion grounding, it requires a sound investigation and understanding of the reasons for the inclusion of those particular “emotion primitives” in the architecture (in terms of their functions) and of the mechanisms for their integration with other elements of the architecture (in terms of their integration with different cognitive and behavioral subsystems). Within HUMAINE, this issue will be approached jointly with WP3 by developing guidelines for the operationalization of theoretical emotion models into computational ones.

Two other lines of research, little developed so far, can also greatly contribute to avoid those two “dangers”. The “emergent” approach to emotion modeling, in which behavior that an observer could consider as “emotional” but that arises from the interactions of the system with its environment, rather than from explicit “emotion components” in the architecture, can contribute to avoid the risk of “over-attribution”. It can also improve our understanding of emotional phenomena (and avoid “over-design”, i.e., the elaboration of systems unnecessarily complex) by uncovering some aspects of emotions that can be accounted for by simple cognitive and behavioral mechanisms and their interactions, without the need to postulate specific “emotion machinery”. Emotion grounding can be better achieved by developing computational models of emotions that take developmental and evolutionary perspectives, in

which emotional systems form (or “grow”) in the course of the interactions of the artifact with its (physical and social) environment over the life time of an individual (development) or a “species” (evolution). In these cases, emotions would acquire a meaning not only for the human designer and user, who would be able to track and understand the reasons behind those particular emotional systems, but also for the artifact itself. Some work will be done within HUMAINE in these two aspects, hopefully bootstrapping further research in these areas in the future.

### **6.2.2 Dissolving the “mind-body” problem**

Related to the grounding problem of emotions in artifacts is the problem of investigating and establishing well-founded links between “higher” and “lower” levels of cognition and action and the influences of emotions in both. As we have seen in previous sections, different conceptual traditions in AI, which can be roughly classified into symbolic and embodied approaches, put the emphasis on these different “higher” and “lower” levels, and Section 5.2.2 has outlined some of the integration challenges that developing comprehensive, “complete” emotion architectures poses. Hybrid architectures that bring together “the best of both worlds” are explored at present, but sound progress towards solving this problem needs to go beyond the type of solutions that current state-of-the-art research allows to provide, and that often reminds us too much of the “pineal gland” that Descartes postulated as a link between “body” and “mind”. Ideally, future systems should not provide “hybrid” solutions but rather a true integration that departs from a dualist stance and dissolves the “mind-body” problem. We are aware of the extreme difficulty of this goal and would not dream to claim to be in a position to achieve it, since this problem has been pervasive in philosophy for centuries and is still an open (and very “hard”) one. However, this goal should be regarded as an ideal to guide research. As can be seen in Section 8, different exemplars in this workpackage will investigate problems underlying the involvement of emotions in “lower-level”, “higher-level”, and “middle-level” (the “gap” between those) cognition and action. Even if the outcomes of these exemplars will probably be in the line of “pineal gland” solutions, their elaboration will shed light on a number of problems that need to be tackled if we want to progress towards more sound solutions. Examples of such problems are:

- The roles that emotions play in the synchronization of numerous cognitive, behavioral and bodily subsystems.
- The mechanisms needed to bridge the gap between the “internal” and “external” aspects of emotions in order to synthesize expressive behavior truly grounded in the architecture of the artifact (see Cañamero and Gaussier, 2004, for a discussion of this latter problem).
- The integration of multiple levels of emotion generation.

### **6.2.3 Untangling the “knot of cognition”: the links between emotion and intelligence**

The surge of research in computational models of emotions and affective computing in general is often conceptualized as a consequence of a change in paradigm regarding the role of emotions in intelligence. Overcoming the tradition that regarded emotions as undesirable consequences of our embodiment that impaired reasoning and decision making, nowadays emotions are considered as pervasive in many aspects of cognition and action and an essential

element of intelligence. Even if we agree with the latter position, this view should not become an unquestioned assumption, as this would give rise to rather superficial computational models of emotions and poor understanding of our achievements. On the contrary, sound progress in the area necessitates “dissecting” or “untangling” this “Gordian knot” by investigating and modeling the mechanisms underlying different aspects of the involvement of emotions in cognition and action, and carefully assessing which aspects, among the many possible, can be meaningfully modeled in our artifacts and which ones are specific to human or animal emotional systems, and therefore meaningless in the case of artifacts. Mechanisms underlying emotional modulation of different aspects of cognition and action need to be singled out, in close collaboration with WP3, and implemented in computational models. This is not sufficient, however, as emotions are not “isolated” or “independent” entities or modules within the brain or the artifact architecture, but are deeply intertwined and rely on many other aspects and subsystems; therefore, these elements must be investigated and modeled in parallel with emotions. Some of the relevant notions are, for example: the notion of “self”, of which only some rudiments around the ideas of “bodily self” (e.g., proprioceptive feedback) and “autobiographic self” (e.g., autobiographic and emotion-related memory) can be meaningfully implemented in artifacts; mechanisms for social motivation; or mechanisms underlying simple forms of “empathy” (e.g., mechanisms of emotional contagion such as sensory-motor coordination and synchronization, mimicry, feedback, etc.). One of our exemplars, developed in close collaboration with WP3, will be devoted explicitly to the investigation of mechanisms underlying the link between emotions and different (low-level) aspects of cognition and action, and other exemplars will include work on some of these other issues.

#### **6.2.4 Measuring progress: Which are the contributions of emotions to our systems?**

Finally, an important requisite for the advancement of the area is the ability to assess the benefits that the inclusion of emotional systems brings to our artifacts. The fact that emotions seem to fulfill a number of important functions in humans and provide increased complexity and flexibility of behavior in natural systems cannot be used as a justification of the value of artificial emotional systems. The inclusion of emotional element in our architecture does not make our artifact more valuable *per se*. On the contrary, we must be able to show accurately and precisely that (or rather whether) our results allow us to conclude that emotions improved the performance or the interaction capabilities of our artifact and how. An obvious way of doing this is by running control experiments in which the artifact performs the same task “with” and “without” emotions, and comparing the results. Although some criteria and indicators of performance have been proposed, much research is still needed to develop different criteria, performance indicators, scenarios, testbeds, etc., to achieve not only qualitative but also quantitative evaluations of our artifacts adapted to different types of applications and problems. To take some initial steps in this direction, the various exemplars will devote efforts to this problem, and also interaction with WP9 (Usability) will provide very valuable feedback.

## 7 Relation to other workpackages

The task of modeling the roles of emotion in cognition and action calls for distinctive (inter-disciplinary) expertise within WP7. However, its relations with the problems addressed by the other thematic areas should not be overlooked, and appropriate interfaces with them will be implemented throughout the lifetime of the project. The particular links that will be fostered will depend on the exemplars finally selected for development. Potential cross-links have been pointed out throughout the deliverable, as currently foreseen, and will not be systematically summarized here. This section will rather outline some points in which collaboration with other WPs would be particularly beneficial, as seen from the perspective of WP7.

### 7.1 WP 3: Theories and models

The links with this workpackage are very strong, as pointed out at numerous places in the deliverable, since these workpackages provide complementary approaches (analytic in WP3, synthetic in WP7) to the study of the involvement of emotions in cognition and action. With the goal of bringing together these approaches in a way that fosters mutual enrichment and a deeper understanding of this thematic area, some of the exemplars proposed will be developed jointly by both workpackages (cf. Sections 8.1 and 8.2). Expected benefits of this joint effort include:

- The development of guidelines for the operationalization of theoretical models in computational implementations. This endeavor should bring clarity and preciseness to theoretical models and the notions they use, often too vague or ambiguous; at the same time, it should bring soundness to computational models, often too ‘ad hoc’ as a result of a superficial understanding of the complexity and implications of their underlying theories and concepts.
- The development of biologically / psychologically plausible emotion-oriented artifacts that at the same time: (a) constitute testbeds for an improved understanding, experimentation and testing of emotion theories in ways that cannot be done with humans and animals, due either to the complexity of the task, or to ethical considerations; and (b) provide more sound and plausible solutions to problems in human-computer interaction and autonomous agents research.

### 7.2 WP 4: Signals to signs of emotions

Understanding how emotionally “neutral” signals are transformed into (or interpreted as) emotionally significant signs and signs of emotions is a key problem in the relation between emotion and cognition (elicitation of emotions, appraisal). The goal of WP4, the generation of emotional signs through processing and analysis of multimodal input signals obtained from users, can therefore shed much light to this problem. Contributions that links with WP4 could make to WP7 include:

- Provide guiding principles regarding features that the architectures studied and developed under WP7 could use to process and recognize emotionally significant perceptual “input”, as well as those that our artifacts could produce in their behavior to express their

emotional states in a way that human users can meaningfully interpret as signs of emotions.

- Provide guiding principles regarding mappings between signals and signs of emotions that user models should integrate to model, recognize and keep track of the emotional state of the user.

At the same time, the artifacts studied and developed in WP7 can provide new challenges to WP4, since their perceptual, cognitive and behavioral (expressive) mechanisms are very different from those of humans (think of the sensors and actuators of robots, for example) and, nevertheless, must also transform signals into signs of emotions that can be understood by humans if they have to be autonomous and proactive partners in their interactions with them.

### 7.3 WP 5: Data and databases

Relations with this workpackage would center around issues regarding validation of computational models. These issues were some of the core topics recently discussed at the 2004 Spring Symposium of the American Association for Artificial Intelligence on “Architectures for Modeling Emotions: Cross-Disciplinary Foundations” (Hudlicka and Cañamero, 2004). Some of the questions that were debated in this respect include:

- Availability of empirical data:
  - What empirical data are required to generate computational models at different levels of granularity, and what is their availability?
  - Does neuroscience provide adequate data to allow modeling these phenomena at the circuit and neuroanatomical components?
  - Does psychology (and neuroscience) provide adequate data supporting empirically-motivated choices for representation and inferencing involving internal mental constructs and processes (e.g., goals, expectations, plans)?
- Model and architecture validation:
  - How are the validation criteria and metrics different for different levels of granularity? For different types of architectures? For different phenomena modeled?
  - What are the best ways of coupling computational modeling and empirical research approaches, for the purposes of data generation and hypothesis validation?

The elaboration of appropriate (sound, representative and tractable) databases certainly appears as a necessary step towards answering these questions, and provide the means to undertake much-needed cross-disciplinary validation of computational models.

### 7.4 WP 6: Emotion in interaction

The links with this workpackage are also very strong, as indeed it is very difficult to establish clear-cut boundaries between “action” and “interaction”, and there is again room for the joint development of an exemplar (cf. Section 8.4). Points in which collaboration would be mutually enriching include:

- Investigation of different cognitive and behavioral mechanisms underlying the role of emotions in social interaction, e.g., low-level, automatic mechanisms such as emotional contagion, attention and memory in social interaction, the generation of socially meaningful behaviors, or mechanisms underlying different aspects of social dynamics.
- Study of mechanisms underlying the link between behavioral expression of emotions and the underlying emotion synthesis systems.
- Elaboration of models of the emotional state of the user for human-machine interaction in a way that allows to anticipate her/his actions, as well as the eventual reactions of the system.

## 7.5 WP 8: Emotion in communication and persuasion

Relations with this workpackage are principally foreseen regarding cognitive and behavioral mechanisms underlying communication and persuasion. Potential areas of collaboration with WP8 include:

- The use of BDI (beliefs, desires and intentions) architectures in communication and persuasion.
- The role of user models in communication and persuasion.
- Embodied conversational agents as platforms to investigate the role of emotions conveyed via different communication modalities in persuasion.
- Architectural (emotional, cognitive and behavioral) elements required for emotion synthesis that backs up the use of expressive behavior in communication and persuasion.
- The emergence of deceptive and persuasive behaviors as a result of the (co-)evolution of emotional expression and cognitive and behavioral mechanisms that allow the recognition of and response to emotional expression (e.g., simulated in artificial life environments).

## 7.6 WP 9: Usability

Like WPs 3 and 10, this workpackage is somewhat orthogonal to the other workpackages. Some particularly relevant links with WP9 include:

- Requirements for long-term interactions between human users and emotion-oriented systems. As mentioned in Section 2.1, emotion architectures underlying the behavior of emotion-oriented systems endow these systems with the features of autonomy and coherence necessary to sustain believability in long-term interactions. Collaborative work would investigate the requirements for architectures to achieve this goal in different kinds of applications and interaction scenarios.
- Acceptability: guidelines on what emotions would and would not be accepted by users (and therefore should or should not be generated by the underlying emotion architecture) in different types of interactions.

- Cultural and individual differences: current computational models of emotions have much difficulty integrating these aspects, due partly to problems in those models and partly to the variety and complexity of those differences. WP9 could provide guidelines regarding which aspects computational models should attempt to take into account for different types of interaction scenarios.
- Evaluation methods: The acceptability and usability of emotion-oriented systems depends to a big extent on the adequacy of the underlying synthesis system that generates its “emotional states”. The evaluation of the architecture should thus take into account, and possibly benefit from, the methods used to assess the usability of the system. Conversely, methods used to analyze the performance of emotion architectures can provide contribute to improve usability methods, in particular ethological (quantitative) analysis of the behavior and interactions of the emotion-oriented system.

## 7.7 WP 10: Ethics and good practice

Potential relations with this workpackage are, again, numerous. To outline a few:

- Regarding ethics:
  - Elaboration of guidelines concerning features and data about human users to be included in user models.
  - Analysis of the ethical implications of endowing artifacts with emotional systems, i.e., systems that “have” (some sort of) emotions and, most importantly, that are perceived by human users as having emotions and feelings (over-attribution problems).
- Regarding good practice:
  - Substantiated assessments of the current capabilities, properties, and scope of implementation models of the link between emotions and cognition and action.
  - Implications of the current understanding of how affect and emotion influence and determine action selection and behavior at temporally local and global, and at socially individual and group levels.

## 8 Preliminary ideas about possible exemplars

This section sketches some preliminary ideas regarding possible exemplars that could be meaningfully developed within WP7. It should be stressed that these are initial ideas and not definitive plans; they illustrate the types of exemplars that could very likely be developed within the Workpackage, based on initial proposals and discussions among partners; therefore, some differences are to be expected between these preliminary plans and the exemplars actually developed.

Ideally we would like to investigate as many of the issues listed below (within the time and resources available in the project), rather than selecting one of them. Work on these different “exemplars” will be carried out by moderately-sized working groups, to ensure optimal cooperation. The initial suggestion is to develop these “exemplars” as part of the common theme of “emotions in decision making” (including action selection), since decision making involves a wide range of cognitive and behavioral capabilities in which emotions are involved. All the different exemplars should include a critical analysis and review of existing work and work developed. To foster critical analysis and understanding of emotion models, exemplars 8.1 and 8.2 would be developed jointly by WP7 and WP3, with WP3 focusing on the biological and psychological plausibility, along with the computational adequacy and coherence of the exemplars, in regard to the actual developments of (appraisal) emotion theories. Ideally, critical analysis would also involve a philosophical reflection that puts in perspective research on computational models of emotions and cognition and action; this could be a collaboration between KCL (Peter Goldie et al.), UH (Lola Cañamero), GERG (Klaus Scherer) and OFAI (Paolo Petta).

### 8.1 First Possible Exemplar: Emotion in “lower-level” cognition and action

*Participants: UH, GERG, OFAI*

This exemplar is to be developed jointly between WP7 and WP3: the theoretical aspects of critical analysis of the underlying emotion models and the elaboration of guidelines for operationalization of emotion theories will be investigated within WP3, whereas the computational models and their comparison will be the task of WP7. This exemplar would investigate what the literature refers to as the “fast pathway” of emotion processing (LeDoux, 1996; Rolls, 1999). Among the many different problems that could be investigated here, we propose to focus on the following ones.

#### 8.1.1 Mechanisms underlying emotional modulation of cognition and action

This exemplar would be based on the ideas presented in Section 5 regarding this problem, and could be the outcome of a collaboration between UH (Lola Cañamero and Richard French), OFAI (Paolo Petta, Bernhard Jung, Erik Hörtnagl) and GERG (Klaus Scherer, Etienne Roesch, David Sander, Didier Grandjean). It will investigate mechanisms underlying the influence of emotions in (“lower-level”) cognition and action at various levels.

The research project of Richard French as UH postdoctoral fellow for WP 7 will take Fellous neuromodulatory model of emotions (Fellous, 1999, 2004) as inspiration. It will explore different ways in which emotions, via neuromodulation of underlying “neural substrates”, can affect different aspects of (“low-level”) cognition and action in various control (action selection) architectures for autonomous robots with increasing levels of complexity. Different action selection architectures with increasing levels of cognitive and behavioral complexity (and therefore increasing “flexibility” and potential for neuromodulation) will be defined in the course of the project. The architectures will be designed taking inspiration from those proposed by Valentino Braitenberg in his “Vehicles” (Braitenberg, 1984), and could for example be:

1. A 7-neuron version of the type 2 Braitenberg vehicle (phototaxis) that changes from light-avoiding to light-seeking behavior by modulation of its PDP interneuron transfer functions.
2. A variant on 1 that includes the ability to signal to other vehicles the onset of a change in its behavior. This vehicle also includes the ability to receive and so be affected by the same signal broadcast by a conspecific.
3. A variant on 2 that includes the ability to sensitize the vehicle’s sensory system and so make it more sensitive to stimuli from its environment.
4. A variant on 3 that uses the notion of ‘Primers’ and ‘Releasers’ to choose between a larger range of competing behaviors in a ‘behavior system’.
5. A variant on 4 that uses the notion of a ‘Value system’ to learn useful mixes of modulators to affect its behavior system.
6. A variant on 5 in which the input to the value system is also affected by modulators.
7. In contrast to the above ideas, here we start to use evolution to reduce the amount of human input (and so prejudice) in designing a vehicle.

...and if we had a million Euros ...

8. A variant on 7 that explores neuromodulation as the basis of ontogenetic development for a ‘young’ vehicle.

OFAI has been investigating how emotions can influence cognition and action in action selection and decision making architectures for autonomous agents inspired by other emotion theories, namely appraisal and motivational / adaptational models, such as the TABASCO architecture (Staller and Petta, 1998; Petta, 2003). The sensing interface of situated embodied instantiations of this architecture - in particular to visual modalities (cameras) - has highlighted the role of anticipation and assessment of making and breaking, as well as preserving, coordination of internal states to what the environment offers. Such internal states are configurations of *processes* - dynamic interpretations of situational meaning - that engage according to whether current coupling to the environment supports their execution. Investigations on the affective nature of such *conceptual coordinations* (Clancey 1999, Clancey 2000) and their integration into the architectural design are the focus of OFAI’s participation in a new large national joint research project (FWF 2003) on cognitive vision as key technology for intelligent assistance. The engineering challenges to provide such coordination of bottom-up and top-down control for observational tasks (e.g., in hide and seek scenarios) thus comprise: run-time identification of possible and selection of the most useful interpretation of visual information (utilizing a given instrumentation of detectors, recognizers, trackers, spatio-temporal reasoners, etc.); regulation and control of establishment and termination of commitments to specific interpretations, and of effort invested in corroborating and re-establishing such commitments in the light of partial and uncertain information and intermittent data failures (ranging from sensor aberrations to downright

hiding of a target); management of distributed pre-attentive and attentive processes. With mobile robots, further affective dimensions of active vision will be introduced, such as active identification and selection of (“preference for”) available scenes/locations and their interpretation (Hörtnagl *et al.*, 2004). Taken together, these component technologies will provide important insights into the nature of mechanisms underlying emotional modulation of cognition and action, design constraints and interdependencies, and will also contribute to the exemplar described under Section 8.1.2.

The collaboration between UH and OFAI will investigate the differences / commonalities that these different mechanisms underlying the link between emotions and cognition / action entail in different perceptual / attentional and behavioral tasks in the context of action selection. The participation of GERG will focus on the biological plausibility, along with the computational adequacy and coherence, in regard to the actual developments of the appraisal theories.

### 8.1.2 Emotion elicitors

The study of mechanisms underlying the interrelation between emotion and cognition and action does not necessarily provide insight regarding the mechanisms that elicit emotions. One interesting exemplar could investigate the involvement of appraisal processes in “low level”, “non-symbolic” emotion generation and action selection mechanism. This would be the outcome of a collaboration between GERG, University of Geneva (Klaus R. Scherer, D. Sander, D. Grandjean, and E. Roesch) and ASRG, University of Hertfordshire (L. Cañamero, O. Avila-García, R. te Boekhorst, Richard French). The following idea could be the starting point for this exemplar. The members of the UH group are investigating the problem of action selection in autonomous robots. So far, the PhD work of O. Avila-García has implemented different architectures for action selection using a behavior-based, “nonsymbolic” AI approach. Motivations and Behaviors were used to codify the goals of the system – e.g., to keep an optimum level of energy – and ways to fulfill them – e.g. recharging energy – respectively. In (Avila-García and Cañamero, 2004) they have proposed that, acting at perceptual level – that is, modulating the relevance of the sensory inputs of the robot’s action selection process – the adaptivity of the robot can be enhanced. In a series of experiments, a “hormone-like” feedback mechanism that modulates the relevance given to external stimuli (exteroception) was proposed. In following experiments, it is planned to modulate also internal stimuli (interoception). It appears that appraisal processes would be determinant for the generation of these “hormones”, and that a role of emotion could be to modulate sensory and attentional processing – either for internal or external inputs – to produce adaptive behavior according to different environmental circumstances. However, so far, research developed by the UH group lacks a study of the actual mechanism that might trigger the “hormone generation” as well as the modulation of sensory and attentional processing. An important issue is therefore to identify when and how emotional processes appear and bias the relevance given to different sensory inputs. Scherer and the Geneva Emotion Research Group developed a component process model of emotion (Scherer, 2001) that could complement the implementation described above and be tested in this framework. A critical appraisal mechanism that would represent a fascinating first step for modeling would be “relevance detection” because it is starting to be well studied both at the psychological and at the neural levels (Scherer, 2001; Sander *et al.*, 2003).

Finally, to assess the contribution of this research, the resulting behavior of the robots will be evaluated using, on the one hand, different indicators of performance developed at UH (Avila-García, *et al.*, 2003; Avila-García and Cañamero, 2004), based on the notion of “viability” (Ashby, 1952) of stability of the internal parameters of the robot, and on the other hand

ethological analysis of the observed behavior. Ethologists use a suit of statistical tools (e.g., Markov modeling, cluster- and factor analysis) to reveal temporal patterns in the sequence of behavioral elements. These patterns are then used in the construction of causal explanations for various aspects of motivation-mediated behavioral phenomena such as action selection. Much less attention has been devoted so far in ethology to the explanation of emotion-mediated behavioral phenomena. Such an analysis, however, would put forward several important ideas, such as: (a) Working with robots in the “real world” brings to light “side effects” that would not show up or would go unnoticed in computer simulations; (b) Ethological methods pick up the manifestations of these side effects in a systematic way (although it is not guaranteed that they can also *explain* the behavioral patterns); and (c) Some of these “side effects” are easily (and often unnecessarily) over-interpreted by human observers as reflecting specific “drives”, “motivations” or “emotions”.

This exemplar might show, first, the utility of appraisal processes, including relevance detection, to generate “hormones”, to modulate attentional processing, and to adapt robots’ behavior to different environmental conditions. And secondly, to show that “low level”, “non-symbolic” models of emotion, attention and relevance detection may produce interesting behavioral phenomena that could well be identified with processes in animals and humans. In conclusion, this exemplar would represent a joint effort to implement a model of appraisal processes in autonomous robots that would be of high importance to study emotion generation and action selection.

## 8.2 Second Possible Exemplar: Emotion in “higher-level” cognition and action

*Participants: UNI. DI BARI, France Telecom RD, QUB, University of Geneva (GERG), EPFL*

This exemplar will also be developed in close interaction with WP3. The various aspects of affective states which are examined in the deliverable *D2a: Report on initial plenary meeting* (Table 1 of WP3) are tightly intertwined. Those of them which are classified as having a ‘slow rapidity of change’ (interpersonal stances, preferences/attitudes and affect dispositions) influence activation and intensity of ‘fastly changing’ ones (the emotions). Some slowly changing types are related to values and norms, to personality traits or to social relations. Repeated (in time) emotion mixtures influence moods.

We wish to investigate and represent, in ‘high-level cognition models’, this relationship among affect types, in terms of relationship between beliefs, desires, goals, intentions and emotion activation. We wish to study how emotion activation is influenced by interpersonal stances, attitudes and affect dispositions and by the social context in which the triggering event occurs and is perceived; we wish to study how these influencing factors may be modeled in their turn. Finally, we wish to investigate how the effects of emotions on a revision of beliefs, goal priority, reasoning style and decision making may be modeled in the same framework.

The models we propose to develop may be situated across some of the categories which are mentioned (again) in D2a, WP3: they are ‘appraisal models’ in that they assume that emotions are elicited by a cognitive evaluation of events; they may be seen as ‘motivational models’ as the intensity of activated emotions is a function of motivation and action tendencies; finally, they are, in a way, ‘lexical models’ because the emotions triggered are closely related to the categories of goals involved in the event occurred.

Cognitive models of emotions require going over the old distinction between ‘rational’ and ‘emotional’ thinking, to extend the Belief/Desire/Intention theory and models with emotional factors. This extension increases the need to include uncertainty in the representation of the relationships among the various components: in doing that, we plan to examine advantages and disadvantages of various AI approaches to knowledge representation and reasoning (e.g. classical logic, fuzzy logic, belief networks, Dempster and Shafer’s theory, neural networks...).

As suggested by the Coordinating partner of WP7, after an initial state-of-the-art study we plan to consider a subset of the emotions which might be activated in the scope of a decision-making process: we will go into theories about their activation and effects and will study how represent them in terms of ‘cognitive BDI&E models’.

Another aspect that must be considered here concerns emotional factors in user modeling. The following example provided by EPFL illustrates how this can be done in the context of decision making. The integration of the existing and novel advances concerning models of emotions and the role of emotion in the modulation of action can be exemplified and applied to a real-life situation in the shape of an adaptive interface for control centers. A demo application would feature a multimodal interface to a control center for a transport terminal (metro, train station, airport, etc.) or a similar scenario. The interface will provide different control mechanisms, ranging from “classical” GUI interfaces to have low-level and fine control of each and every parameter of the system being controlled –lights, fans, gates, and so on-, up to a body-gestures and speech-based interface able to react to the user’s commands – arm gestures, speech- in a fast way even if this requires loosing some of the fine-control. The interface will continuously analyze the emotional state of the user and adapt the interface mechanism (paradigm) accordingly; e.g., while in “normal” mode the controller must go through several windows on the screen in order to set the correct parameters of a given device. However, if the system detects the user is under a certain emotional (stress) state, it will automatically switch to a gesture-speech mode allowing for high-level control of the critical parameters. In this mode the system is less flexible, but reacts faster and adapts to the situation.

Within this framework, we envisage two categories of possible outcomes:

- 1) Comparison and testing of different kinds of models in a given situation; this could be for instance a car driving situation, or decisions such as “decide whether to stop smoking”, or “decide whether to accept a difficult health care treatment”;
- 2) Design / specification of a tool which enables simulating how emotion activation and effect vary when several factors regulating the activating conditions are manipulated: characteristics of the event, social context, attitudes, dispositions and so on. This work could be based in an in-depth critical analysis of tools already developed by different partners involved in the exemplar.

### **8.3 Third Possible Exemplar: bridging the gap between “lower-level” and “higher-level” cognition and action**

*Participants: USAL, INESC-ID, IST, OFAI*

This exemplar shall demonstrate the ecological adequacy (in terms of parsimony and coverage) and engineering utility (in terms of desirable design- and runtime qualities) of a

whole (Pfeifer 1996) affective architecture in a strategic dynamic environment with bounded/unbounded levels of indeterminacy, and the appropriateness (the believability) of the agent's responses. A proposed solution approach would involve an evolution of the TABASCO (tractable appraisal-based architecture for situated cognizers) framework developed at OFAI (Staller and Petta, 1998; Petta, 2003) to integrate continuous planning and advanced behavioral components developed at USAL (Aylett and Barnes, 2000, Aylett et al. 2000, Avradinis and Aylett 2003), in which emotional states link to motivations which control multiple interactive planning and execution processes of an agent, while preserving a sound grounding in psychological theories and models provided by WP3, e.g. (Leventhal and Scherer, 1987).

Behavior generation will also have to include coordinating mechanisms of coping and regulation, considering different scopes of current dynamical and situational context (Dias and Paiva, 2004). On the other end, perceptual aspects such as attention will filter, emphasize or suppress information from the external world (Martinho and Paiva, 2003).

The exemplar could be demonstrated using a graphical character in a virtual world.

Low-level physiologically-oriented / pre-attentive accounts of emotion integrate well with behavioral architectures in embodied agents, but these architectures are reactive and therefore very subject to the problems of local decision-making. They are also more appropriate to non-human animals in which reactive behavior appears more dominant and in which language actions are not part of the action repertoire. On the other hand, high-level appraisal-based accounts of emotion integrate well with symbolic AI architectures in which action-selection involves planning and other mechanisms supporting the construction of richer expectations and action sequences moving into the future. However, such architectures have known problems with sensor-directed processing and environment-dependent multi-tasking and are thus often brittle and relatively impervious to change in the world. This exemplar seeks to integrate high and low-level architectures using a model of affect as part of the integrative mechanism.

Continuous (or continual) planning refers to systems in which planning and execution are indefinitely inter-leaved and new goals are generated and managed over time. One mechanism for handling the generation, reordering and replacement of goals is that of molar and molecular motivational constructs from cognitive psychology such as concerns (Frijda, 1986) and core relational themes (Smith and Lazarus, 1993; Lazarus, 2001). A possible design for such a "motivation-driven continuous planner architecture" can be seen in Figure 1. In this diagram, motivations are involved in the generation of goals and could also be involved in the selection of goals to plan for or actions to execute. Executable actions can be produced as contextual switching for a low-level architecture in which a lower-level emotion system both interacts with sensing and with low-level action-selection. Meanwhile motivations can be used to combine a high-level appraisal-based system interacting with goals and the low-level affective systems from the execution systems. This exemplar would investigate these interactions and dependencies and generate test scenarios in which the two types of affective system and the two levels of actions selection can be evaluated.

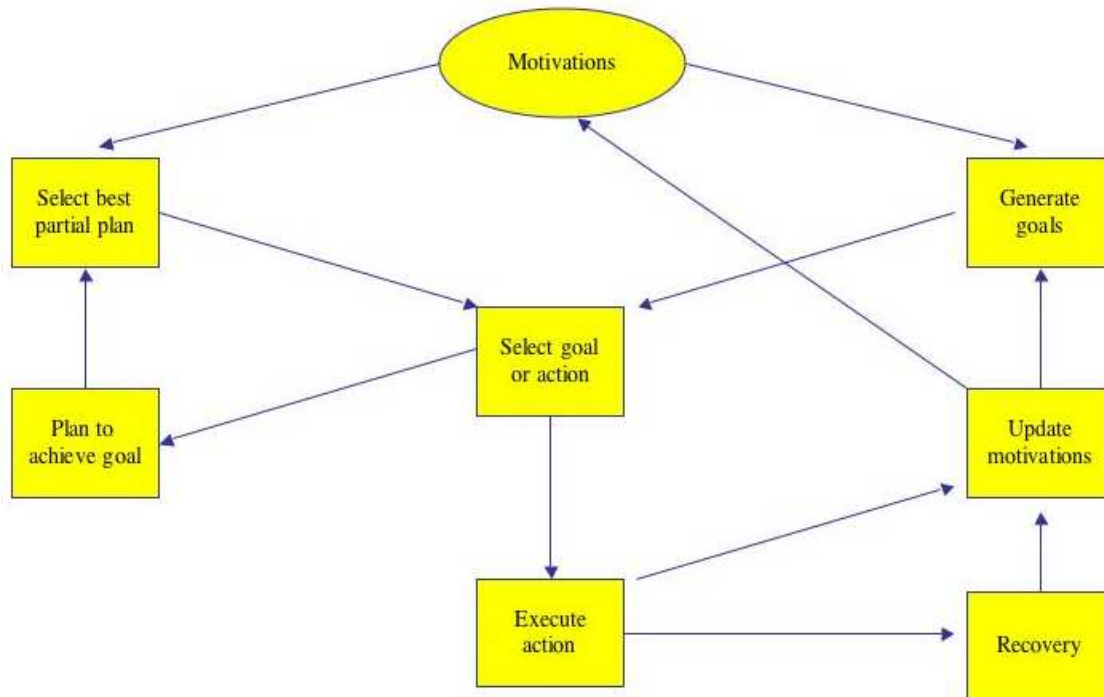


Figure 1: a motivation-driven continuous planner design

Given the final goal of integrating the different parts of an architecture needed for believable action in a complex world, a first step will be to identify possible decompositions of necessary functional elements of such an agent, relating it to theory as developed in WP3 and technologies available at the partners involved in the exemplar work. The features of such a decomposition, including the communications between the parts; the data each of the components uses; the inter-dependencies; the timing behavior and synchronization issues; taken together with the characteristics of the system they are embedded in – the agent body and its simulated physical coupling to the environment – are to provide a set of clear-cut criteria for the evaluation and comparison to existing affective architectures, including such that cover only part of the raw sketch presented here.

## 8.4 Fourth Possible Exemplar: Emotions in Social Cognition and Interaction

*Participants: Paris8, UH, DFKI, MIRALab, OFAI*

This exemplar would investigate various mechanisms underlying emotions in social cognition and interaction. Due to the variety and diversity of topics covered and of platforms used (virtual human agents, ECAs, robots and artificial life simulations), output produced by this exemplar would take the form of critical analysis and common reflection regarding the integration of the different aspects investigated, rather than an implementation or a well-defined design. This exemplar establishes a strong link between WP7 and WPs 6, 8, and 9, and some aspects of it might be developed jointly across these Workpackages.

### *Automatic and low-level mechanisms*

Although our subjective experience might lead us to believe that social interaction strongly relies on cognitive evaluation and modeling of the other partners involved, their emotional

expressions and states, the interaction, etc., many aspects of social interaction and of emotion “processing” in this context are very automatic, relying on low-level (cognitive and behavioral) mechanisms. A particularly relevant example is our ability to “catch” other people’s emotions without been aware of it – the phenomenon known as *emotional contagion* (Hatfield *et al.*, 1994), defined as:

“the tendency to automatically mimic and synchronize facial expressions, vocalizations, postures and movements with those of another person and, consequently, to converge emotionally.” (Hatfield *et al.*, 1992, pages 153-154)

Emotional contagion is important in social interaction as it promotes behavioral synchrony and the continuous tracking of the emotions of others. Conscious cognitive process are involved in more complex forms of contagion, imitation and empathy, but the simplest, more primitive forms of emotional contagion are based on very low-level and automatic processes such as mimicry, synchrony, feedback, and ANS coordination. At UH, the PhD work of Arnaud Blanchard (working with Lola Cañamero) takes a “Perception-Action Model” (Preston and de Waal, 2002) to investigate these phenomena by modeling them in small-sized mobile robots that have to interact among themselves to accomplish a task, and possibly also interact with humans.

### **Social attention and memory**

Paris 8 (Catherine Pelachaud and Christopher Peters, HUMAINE postdoc) will develop autonomous, socially compatible, virtual human agents/ECAs, work that makes contributions to several WPs (6, 7, 8, 9). Within the exemplar of WP 7, this group will investigate the influence of emotions in various aspects of cognition in the context of social interaction, namely attention and memory.

*Attention* (visual attention towards faces as part of a general environmental attention mechanism for early social interaction): Faces and emotional expressions seem to carry special significance for our visual system. Visual attention to faces can be fairly automatic and mandatory in nature. Emotional expression in a face can be perceived outside of the focus of attention and can guide focal attention to the location of the face. It is important to pay attention to faces for recognition and possible social contact to take place among agents. This theme considers an agent immersed in an environment before any social contact has taken place, and considers how social contacts may start. For example, animations featuring social interactions often show multiple agents communicating, but how did the agents meet in the first place? In order to model this, agents should pay special attention to other agents in their environment, particularly their faces, for recognition and interaction to take place. However, this sort of social attention should also be compatible with other forms of attention.

*Memory*: Memory would appear to form an important basis for agent behavior during social interaction. For example, an agent’s memory of social status and relationship with other agents may effect the amount of attention they receive. Furthermore, memory and emotion are intertwined; accessing emotional memories may cause a change in the agents current emotional state. In terms of agent memory, there would appear to be three important areas of agent long-term memory open for consideration: (1) Autobiographical memory of personal events and experiences that shape who we are. Self-description, emotional experiences, and personal experience of events are important for constructing and storing the agents personality. (2) Social memory of perceived personality traits and identification of others, perceived social relationship network involving others. Cognitive basis for social perception. Organization of person impressions. Provides the agent with a memory of social contacts that

may change the attitude and behavior of the agent towards them. (3) Semantic memory for associating interaction concepts, emotion, personalities, objects.

### ***Generating socially meaningful emotional behavior***

MIRALab is very interested in the effect of emotion on action and how these effects can be modeled so that they are correctly reflected by virtual people in a virtual environment, and understood by real humans. Within this exemplar, MIRALab (Nadia Magnenat-Thalmann, Arjan Egges) would thus investigate the issue of generating emotional behavior that can be easily understood by humans. In particular, using 3D graphics, this group would develop and implement models that describe the relationship between cognitive processes and their expression through body behavior, namely:

- Standards for translating a cognitive process into body behavior
- Definition of the basic 'units' that control body behavior
- Models that describe the general mechanism of interaction between cognitive processes and body behavior

Especially interesting can be the effect of physical body properties on motion (shifting balance, eye blinking, etc.) and how these motions are linked with emotion and personality. These issues will be investigated taking advantage of animation systems developed within the group that allow easy parameterization of body behaviors (posture, gesture, etc.) and that result in realistic body animations, paying particular attention to:

- Automatic animation of idle motions (balance shifting, small posture variation)
- Animation and blending of gestures, idle motions and other animation clips (such as walking, sitting, etc.)

### ***Moods and appraisal in social interaction***

At DFKI, the PhD work of Patrick Gebhard is developing a theory-based appraisal model, generating both emotions and moods based on a configurable, subjective appraisal component, in the context of multiple animated conversational characters. The use of multiple animated conversational characters offer more sophisticated dialogue techniques in application areas such as virtual learning environments, storytelling systems, and e-commerce applications: They enrich the repertoire of modalities to convey information and they can serve as a rhetorical device to reinforce beliefs. Moreover, the use of multiple emotional characters allow conveying social aspects such as interpersonal relationships. All together can greatly enhance quality of such systems. But designing believable behavior can become a complex task, especially if user interaction has to be taken into account. Characters must respond in a way both appropriate and non-repetitive because otherwise their believability is undermined. One approach is to enhance existing computational models of emotion, like the OCC model of emotions by social aspects to cover the addressed issues. The design of this model should consider the requirements for input and output interfaces of multi-character applications. Such applications are mostly action/event driven. To compute social aspects it is necessary to identify relevant input for them. Mostly relevant candidates are dialog moves like *insult*, *encourage*, *tease*, etc. These must be subjectively appraised by each character in order to obtain basic emotions and to compute social aspects. Moreover, this approach allows the computation based on the user's dialog contributions of user's emotions. Taken in

consideration that emotions are not lasting very long, which means that they can only be used for the modeling of short term behavior there must be a longer lasting affective state, let's call it *mood*, which can be used for long term behavior. All these information of each single emotional character and the user can be used to compute an overall mood (moral) vector of the group, a social integrity vector, mood extremes, and sub groups with the same mood configuration. This will enable a new level of interaction design, such as problem coping strategies, and motivation strategies for learning environments. From a technical perspective, the implementation of the computational model of social aspect can be seen as a self-contained application, which can be reused for several applications with emotional conversational characters. It is explicitly designed to be used by autonomous and centrally controlled agent systems. The model respectively its implementation allows to specify for each character separately the personality settings, the appraisal of events, actions, objects, and the impact of emotions on the actual mood computation. As mentioned above the model computes a character's emotion, it's mood state and computes an overall mood (moral) vector of the group, a social integrity vector, and other group related specific mood configurations. All input and output information are represented in a XML style language.

Although due to lack of resources this working piece of software could not be substantially modified for HUMAINE purposes, it would be made available within the project for the investigation of the above-mentioned phenomena.

#### ***Emotions in social dynamics: 'micro' and 'macro' levels***

At UH, the PhD work of Susan Attwood (working with Lola Cañamero) is investigating differences in social dynamics generated by simple affect-based interactions among simulated primates, using an Artificial Life simulation. In particular, this work explores to what extent various environmental factors (availability, distribution and dynamism of resources) and aspects of social structure (e.g., different degrees and styles of maternal bonding in the initial stages of the group's life span) give rise to different positive and negative interactions among individuals that can contribute to the emergence of groups with different dynamics and structure, such as "equalitarian" versus "hierarchical" groups. Although due to lack of resources this research and software cannot be customized for the purposes of HUMAINE, it can be used as a basis for analysis of the above-mentioned phenomena and their connection with other aspects of emotions in social cognition and action within this exemplar.

OFAI has started to investigate the supporting role of emotions for the sustenance of social norms and thus social cognition and action (Staller and Petta, 2001). Therein, the function of emotions in bridging the micro-macro gap is modeled in terms of the mutual support of norms and emotions. Furthering of this line of research would in particular investigate the dynamic aspects of this system, including the establishing and dissolving of norms and the processes underlying the kinds of distributed processing as in appraisals in groups (Reisenzein, 2001).

## 9 Conclusions and Way Forward

The preliminary plans for exemplars presented in this deliverable are the results of exchanges and discussions among partners that started to take shape during the first plenary meeting of the project in Saarbrücken last March and continued since then, mostly via email and telephone exchanges. They cover a representative sample of key problems that must be addressed within workpackage to set adequate grounds and bootstrap sound progress in this thematic area. They also cover different levels of integration regarding: (a) the expected outputs – ranging from critical conceptual analysis to designs, implementations, and comparison and testing of existing systems and tools; (b) the disciplines involved; and (c) collaboration with other workpackages. Some of them are based on and extend previous collaboration experiences among partners, others propose new and novel collaborations. As previously indicated, we would like to develop as many of the exemplars proposed as the resources available within the network allow. From this point, our next steps will be:

1. To prepare more detailed descriptions of potential exemplars. Ideally, most or all the exemplars proposed here would be further developed, and possibly some new exemplars might arise. These descriptions of potential exemplars will be presented in deliverable D7c.
2. To make firm decisions on the exemplars to be developed. The final exemplars selected will be presented in deliverable D7d.
3. The workshop of this workpackage, planned for Month 19, will provide an opportunity to present and discuss in detail the selected exemplars. It will also provide a forum for the different working groups to meet and start taking actions for the actual realization of the exemplars. The workshop proceedings will constitute deliverable D7a, and should include input not only stemming from the exemplars but also from leading emotion researchers within the Network and in the international community.

## References

### References cited in the text

- Allen S. (1999). Concern Processing in Autonomous Agents, Ph.D. Thesis, Cognitive Science Research Centre, School of Computer Science, University of Birmingham, UK.
- André, E., Klesen, M., Gebhard, P., Allen, S., and Rist, T. (2000). Integrating models of personality and emotions into lifelike characters. In Paiva (2000).
- André, E., T. Rist, S. van Mulken, M. Klesen, and S. Baldes (2000b). The automated design of believable dialogues for animated presentation teams. In Cassell et al. (2000), pp. 220-255.
- Andry, P., Gaussier, P., Moga, S., Banquet, J.P., and Nadel, J. (2001). Learning and Communication in Imitation: An Autonomous Robot Perspective. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 31(5): 431—444.
- Araujo, A. (1994). Memory, Emotions, and Neural Networks: Associative Learning and Memory Recall Influenced by Affective Evaluation and Task Difficulty. PhD thesis, University of Sussex, May 2004.
- Arbib M. A., ed., (2003) *The handbook of brain theory and neural networks* (2nd edition). MIT Press.
- Arkin, R.C. 1998. *Behavior-Based Robotics*. Cambridge, MA: The MIT Press.
- Ashby, W.R. (1952). *Design for a Brain: The Origin of Adaptive Behavior*. London: Chapman and Hall.
- Avila-García, O. and Cañamero, L. 2004. Using Hormonal Feedback to Modulate Action Selection in a Competitive Scenario. In *Proc. 8th Intl. Conference on Simulation of Adaptive Behavior (SAB'04)*. Cambridge, MA: The MIT Press (in press).
- Avila-García, O., Cañamero, L. and te Boekhorst, R. (2003). Analyzing the performance of “winner-take-all” and “voting-based” action selection policies within the two-resources problem. In *Proc. 7<sup>th</sup> European Conference in Artificial Life (ECAL03)*, pp. 733–742. Berlin-Hedelberg: Springer-Verlag.
- Avradinis N., Aylett R. (2003). Agents with no aims: Motivation-driven continuous planning In T. Rist, R. Aylett, D. Ballin (eds.): *Intelligent Virtual Agents, 4th International Workshop (IVA2003)*, pages 269-273, LNAI 2792. Berlin Heidelberg: Springer-Verlag.
- Aylett, R.S. and Barnes, D.P. (2000). Connecting reflection and reaction - a heterogeneous multi-agent architecture. In: K. Dautenhahn (ed.): *Human Cognition and Social Agent technology*, pages 197-224. John Benjamins Publishing co., Advances in Consciousness research Series.
- Aylett, R.S., Coddington, A.M., and Petley, G.J. (2000). Agent-based continuous planning. *19th Workshop of the UK Planning & Scheduling SIG*, pages 1-10.
- Balkenius, C. and Morén, J. (2001). Emotional Learning: A Computational Model of the Amygdala. *Cybernetics and Systems*, 32(6): 611-636.
- Ball, E. (2003). A Bayesian Heart: Computer Recognition and Simulation of Emotion. In Trapp, R., Petta, P. and Payr, S. eds. (2003). *Emotions in Humans and Artifacts*, pages 303-332. Cambridge, MA: The MIT Press.
- Ball, E. and Breese, J. (2000). Emotion and Personality in a Conversational Agent. In Casell et al. (2000), pages 189-219.
- Bargh J.A. and Chartrand T.L. (1999). The Unbearable Automaticity of Being, *American Psychologist*, 54(7): 462-479.

- Bates, J. (1994). The Role of Emotion in Believable Agents. TR CMU-CS-94-136, School of Computer Science, Carnegie Mellon University, Pittsburgh, USA.
- Bates J., Loyall A., Reilly W.S. (1992). Integrating Reactivity, Goals, and Emotion in a Broad Agent, in *Proceedings of the 14th Annual Conference of the Cognitive Science Society*, pages 696-701, Lawrence Erlbaum, New Haven/Hillsdale/Hove.
- Blumberg, B. (1996). Old Tricks, New Dogs: Ethology and Interactive Creatures. PhD thesis, MIT Media Lab, September, 1996.
- Braitenberg, V. (1984). *Vehicles: Experiments in Synthetic Psychology*. Cambridge, MA: The MIT Press.
- Breazeal, C. (2002). *Designing Sociable Robots*. Cambridge, MA: The MIT Press.
- Cañamero, L.D. 1997. Modeling Motivations and Emotions as a Basis for Intelligent Behavior. In W. Lewis Johnson, ed., *Proc. First Intl. Conference on Autonomous Agents*, pages 148–155. New York: The ACM Press.
- Cañamero, L.D., ed. (1998). *Emotional and Intelligent: The Tangled Knot of Cognition. Papers from the 1998 AAAI Fall Symposium*. Menlo Park, CA: AAAI Press.
- Cañamero, L. (2001a). Building Emotional Artifacts in Social Worlds: Challenges and Perspectives. In *Emotional and Intelligent II: The Tangled Knot of Social Cognition. Papers from the 2001 AAAI Fall Symposium*, pages 22-30. Menlo Park, CA: AAAI Press.
- Cañamero, L., ed. (2001b). *Emotional and Intelligent II: The Tangled Knot of Social Cognition. Papers from the 2001 AAAI Fall Symposium*. Menlo Park, CA: AAAI Press.
- Cañamero, L. (2001c). Emotions and Adaptation in Autonomous Agents: A Design Perspective. *Cybernetics and Systems* 32(5): 507-529.
- Cañamero, L.D. (2003). Designing Emotions for Activity Selection in Autonomous Agents. In Trappell et al. (2003), pages 115-148.
- Cañamero, L. and Gaussier, P. (2004). Emotion Understanding: Robots as Tools and Models. In J. Nadel and D. Muir (Eds.), *Emotional Development: Recent Research Advances*, Oxford University Press (in press).
- Cañamero, L. and Petta, P., eds. (2001). *Grounding Emotions in Adaptive Systems*, Vols. I and II; double special issue of *Cybernetics and Systems: An International Journal*, Vol. 32, Nos. 5 and 6.
- Carberry S., Conati C., Rosis F.de, Gymtrasiewicz P., Hudlicka E., Ishizuka M., Lisetti C., Ortony A., Prendinger H., Revelle W. (2002). Merging Cognition and Affect in HCI: Panel Discussion, *Applied Artificial Intelligence*, 16(7-8): 643-670.
- Carofiglio, V. de Rosis, F., and Grassano, R. (2004). Dynamic models of mixed emotion activation. In L Cañamero and R.Aylett, editors, *Animating Expressive Characters for Social Interactions*. John Benjamins, Amsterdam (in press).
- Cassell, J., J. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjálmsón, and H. Yan (1999). Embodiment in conversational interfaces: Rea. In *Proc. CHI'99*, pages 520–527, Pittsburgh, PA, 1999.
- Cassell, J. and Stone, M. (1999). Living hand and mouth. Psychological theories about speech and gestures in interactive dialogue systems. In *AAAI 1999 Fall Symposium on Psychological Models of Communication in Collaborative Systems*.
- Cassell, J., Sullivan, J., Prevost, S., and Churchill, E., eds. (2000). *Embodied Conversational Agents*. Cambridge, MA: The MIT Press.
- Cassell, J., H. Vilhjálmsón, and T. Bickmore (2001). BEAT : the Behavior Expression Animation Toolkit. In *Computer Graphics Proceedings, Annual Conference Series*. ACM SIGGRAPH.
- Clancey W.J. (1999). *Conceptual Coordination : How the Mind Orders Experience in Time*, Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey, London.

- Clancey W.J. (2000). Modeling the Perceptual Component of Conceptual Learning --- A Coordination Perspective, *Cognition, Education, and Communication Technology Symposium, Stockholm, April 2000, Lund University Cognitive Science*.
- Cos-Aguilera, I., Cañamero, L. and Hayes, G. (2003). Learning Object Functionalities in the Context of Behavior Selection. In U. Nehmzow and C. Melhuish, editors, *Proceedings of Towards Intelligent Mobile Robots (TIMR'03): 4th British Conference on Mobile Robotics*. University of the West of England, Bristol, UK, 28—29 August, 2003.
- Damasio, A. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*. New York: Avon Books.
- Damasio, A. (1999). *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*. London: Vintage.
- Davidson R.J. (1999). Neuropsychological perspectives on affective styles and their cognitive consequences. In Dalgleish T. and Power M.(eds.), *Handbook of Cognition and Emotion*, pages 103-123, Wiley, Chichester/London/New York.
- DeCarolus, B., C. Pelachaud, I. Poggi, and F. de Rosis (2001). Behavior planning for a reflexive agent. In *Proc. IJCAI'01*, Seattle, USA, August 2001.
- de Rosis, F. (2002a). Toward merging cognition and affect in HCI, *Applied Artificial Intelligence*, 16(7-8): 487-494.
- de Rosis, F., ed., (2002b) *User Modeling and Adaptation in Affective Computing*. Special Issue of the journal *User Modeling and User-Adapted Interaction*, Vols. 11-4 and 12-1.
- Dias J. and Paiva A. (2004). “Building Emotional Agents for Interactive Storytelling: a pragmatic approach”, Humaine WP3 workshop on Theories and Models, Geneva, June 2004.
- Duy Bui, T., D. Heylen, M. Poel, and A. Nijholt (2003). Generation of facial expressions from emotion using a fuzzy rule based system. In D. Corbett, M. Stumptner and M. Brooks, editors, *Proceedings of 14th Australian Joint Conference on Artificial Intelligence (AI 2001)*, pages 83 – 94. Springer.
- Dyer M.G., Wolf T.C., Korsin M. (1981). Boris - An In-Depth Understander of Narratives, in Drinan A.(ed.), *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI-81)*, 24-28 August 1981, Univ. of British Columbia, Vancouver, B.C., Canada. Menlo Park, CA: AAAI Press.
- Egges, A., Kshirsagar, S., and Magnenat-Thalmann, N. (2004) Generic Personality and Emotion Simulation for Conversational Agents, *Computer Animation and Virtual Worlds*. 15(1): 1-13.
- Elliott C.D. (1992). The Affective Reasoner: A process model of emotions in a multi-agent system, Ph.D. Thesis, Northwestern University, Illinois.
- Elliott C.D. (1997). Affective Reasoner: personality models for automated tutoring systems, *Proc. Workshop on Pedagogical Agents, 8th World Conference on Artificial Intelligence in Education (AI&ED-97)*, Kobe, Japan.
- Elliott C.D. (1997b). Hunting for the Holy Grail with “emotionally intelligent” virtual actors, *ACM Intelligence*, 1(1).
- El-Nasr, Yen, J. and Ioerger, T. (2003). FLAME - fuzzy logic adaptive model of emotions. *International Journal of Autonomous Agents and Multi-Agent Systems*, 3(3): 1–39.
- Ellsworth P.C., Scherer K.R. (2003). Appraisal Processes in Emotion, in Davidson R.J., et al.(eds.), *Handbook of Affective Sciences*, pages 572-595 (Chapter 29), Oxford University Press, Oxford New York.
- Fellous, J.-M. (1999). The Neuromodulatory Basis of Emotion. *The Neuroscientist*, 5: 283-294.
- Fellous, J.-M. (2004). From Human Emotions to Robot Emotions. In E. Hudlicka and L. Cañamero, eds., *Architectures for Modeling Emotions: Cross-Disciplinary Foundations. Papers from the 2004 AAAI Spring Symposium*, pages 37-47. Menlo Park, CA: AAAI Press.

- Fellous, J.-M. and Arbib, M.A., eds. (2004). *Who Needs Emotions? The Brain Meets The Robot*. Oxford University Press.
- Frijda N.H. (1986) *The Emotions*. Paris: Cambridge University Press and Editions de la Maison des Sciences de l'Homme.
- Frijda N.H. (1993). The Place of Appraisal in Emotion, *Cognition & Emotion*, 7(3&4): 357-388.
- Frijda N.H. (1995). Emotions in Robots, in Roitblat H.L. and Meyer J.-A.(eds.), *Comparative Approaches to Cognitive Science*, pages 501-516. Cambridge, MA: MIT Press/Bradford Books.
- Frijda N.H., Mesquita B., Sonnemans J., Goozen S.van (1991). The Duration of Affective Phenomena or Emotions, Sentiments and Passions, In Strongman K.T. (ed.), *International Review of Studies on Emotion* (Vol. 1), pages 187-225, Wiley, Chichester.
- Frijda N.H., Moffat D. (1994). Modeling Emotion, *Cognitive Studies*, 1(2): 5-15.
- FWF (2003), Joint Research Project "Cognitive Vision – key technology for personal assistance, Forschungsschwerpunkt 91, 15.12.2003-15.12.2006.
- Gadanhó, S.C. and Hallam, J. (2001). Emotion-Triggered Learning in Autonomous Robot Control. *Cybernetics and Systems* 32(5): 531-559.
- Gebhard P., Kipp M., Klesen M., Rist T. (2003). Adding the Emotional Dimension to Scripting Character Dialogues, in Rist T., et al.(eds.), *Intelligent Virtual Agents*. Berlin Heidelberg: Springer-Verlag.
- Gratch J. (1999). Why You Should Buy an Emotional Planner, in Velasquez J.D.(ed.), *Workshop: "Emotion-Based Agent Architectures" (EBAA'99)*, pages 53-60, Third International Conference on Autonomous Agents (Agents '99), Seattle, WA, USA.
- Gratch J. (2000). Socially Situated Planning, AAAI Fall Symposium on Socially Intelligent Agents - The Human in the Loop, North Falmouth, MA.
- Gratch J., Rickel J., Andre E., Cassell J., Petajan E., Badler N. (2002). Creating Interactive Virtual Humans: Some Assembly Required, in Interactive Entertainment, *IEEE Intelligent Systems*, 17(4):54-63.
- Hatfield, E., Cacioppo, J.T. and Rapson, R.L. (1992). Emotional Contagion. In M.S. Clark (ed.), *Emotion and Social Behavior*, special issue of the *Review of Personality and Social Psychology*, 14: 151-177.
- Hatfield, E., Cacioppo, J.T. and Rapson, R.L. (1994). *Emotional Contagion*. Cambridge University Press.
- Heylighen F. (1989). Self-Organization, Emergence and the Architecture of Complexity. In *Proceedings of the 1<sup>st</sup> European Conference on System Science, (AFCET)*, pages 23-32.
- Higgins E.T. (1990). Personality, Social Psychology, and Person-Situation Relations: Standards and Knowledge Activation as a Common Language. In Pervin L.A.(ed.), *Handbook of Personality: Theory and Research*, pages 301-338, Guilford Press, New York/London.
- Hörtnagl E., Pölz P.M., Prem E. (2004). Anchoring Symbols by Mapping Sequences of Distance Measurements: Experimental Results, AAAI Workshop "Anchoring Symbols to Sensor Data", July 25-29, San José, CA, USA.
- Hudlicka, E. (2001). Modeling Affect Regulation and Induction. In Cañamero, L. (2001b), pages 51-58.
- Hudlicka, E. (2004). Two Sides of Appraisal: Implementing Appraisal and Its Consequences within a Cognitive Architecture. In Hudlicka, E. and Cañamero, L. (2004), pages 70-76.
- Hudlicka, E. and Cañamero, L., eds. (2004). *Architectures for Modeling Emotion: Cross-Disciplinary Foundations. Papers from the 2004 AAAI Spring Symposium*. TR SS-04-02. Menlo Park, CA: AAAI Press.

- Izard, C.E. Four Systems for Emotion Activation: Cognitive and Noncognitive Processes. *Psychological Review*, 100(1): 68-90.
- Johns, M. and Silverman, B.G. (2001). How emotions and personality effect the utility of alternative decisions: a terrorist target selection case study. In *Proc. Tenth Conference On Computer Generated Forces and Behavioral Representation*, May 2001.
- Kaiser, S. & Wehrle, T. (2001). Facial expressions as indicator of appraisal processes. In K. R. Scherer, A. Schorr, & T. Johnstone *Appraisal processes in emotion: Theory, methods, research*, pages 285-300, New York: Oxford University Press.
- Kappas A. (2001) A Metaphor Is a Metaphor Is a Metaphor: Exorcising the Homunculus from Appraisal Theory. In Scherer K.R., et al. (eds.), *Appraisal Processes in Emotion: Theory, Methods, Research*, pages 157-172, Oxford University Press, Oxford New York.
- Keltner D., Gross J.J. (1999). Functional Accounts of Emotions, *Cognition and Emotion*, 13(5): 467-480.
- Kravitz, E.A. (1988). Hormonal Control of Behavior: Amines and the Biasing of Behavioral Output in Lobsters. *Science*, 241 (September 30): 1175-1781.
- Kshirsagar S. and Magnenat-Thalmann, N. (2002). A multilayer personality model. In *Proc. 2nd International Symposium on Smart Graphics*, pages 107-115. ACM Press.
- LeDoux, J.E. (1989). Cognitive-emotional interactions in the brain, *Cognition and Emotion*, 3: 267-289.
- LeDoux, J. (1996) *The Emotional Brain*. New York: Simon & Schuster.
- Lester, J.C., S.G. Stuart, C.B. Callaway, J.L. Voerman, and P.J. Fitzgerald (2000). Deictic and emotive communication in animated pedagogical agents. In Casell et al. (2000), pages 123-254.
- Leventhal H. and Scherer K.R. (1987). The Relationship of Emotion to Cognition: A Functional Approach to a Semantic Controversy, *Cognition and Emotion*, 1(1): 3-28.
- Maes, P. (1995). Modeling Adaptive Autonomous Agents. In C.G. Langton., Ed., *Artificial Life: An Overview*, pages 135-162. Cambridge, MA: The MIT Press.
- Marsella S. and Gratch, J. (2002). A step towards irrationality: Using emotion to change belief. In *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'02)*, Bologna, Italy, July 2002.
- Marsella, S. and Gratch, J. (2003). Modeling coping behavior in virtual humans: Don't worry, be happy. In *Proceedings of the 2nd International Conference on Autonomous Agents and Multiagent Systems*. ACM Press.
- Marsella, S., W.L. Johnson, and K. LaBore (2000). Interactive pedagogical drama. In *Proceedings 4th International Conference on Autonomous Agents*. ACM Press.
- Martinho, C., Gomes, M. and Paiva, A. (2003). Synthetic Emotion In: T. Rist, R. Aylett, D. Ballin (eds.): *Intelligent Virtual Agents*, 4th International Workshop (IVA2003), LNAI 2792 Springer-Verlag Berlin Heidelberg.
- Martinho C., Paiva A. (1999). "Underwater Love": Building Tristão and Isolda's Personalities, in Wooldridge M.J. and Veloso M. (eds.), *Artificial Intelligence Today*, pages 269-296, Lecture Notes in Artificial Intelligence 1600, Springer-Verlag Berlin/Heidelberg/New York/Tokyo, 1999.
- Minsky, M. (1986). *The Society of Mind*. New York: Simon and Schuster.
- Moffat D. (1997). Personality Parameters and Programs. In (Trappl & Petta 1997), pages 120-165.
- Morén, J. (2002). Emotion and Learning: A Computational Model of the Amygdala. PhD thesis, *Lund University Cognitive Studies* 93.
- Nilsson, N.J. (1998). *Artificial Intelligence: A new Synthesis*. Morgan Kaufmann Publishers.

- Ortony A. (2003). On Making Believable Emotional Agents Believable. In Trappl R., et al. (2003), pages 189-212.
- Paiva, A. ed., (2000). *Affective Interactions: Towards a New Generation of Computer Interfaces*. Berlin-Heidelberg: Springer-Verlag, LNAI.
- Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford University Press.
- Pelachaud, C., V. Carofiglio, B. De Carolis, and F. de Rosis (2002). Embodied contextual agent in information delivering application. In *First International Joint Conference on Autonomous Agents & Multi-Agent Systems (AAMAS)*, Bologna, Italy, July 2002.
- Petta P. (1999). Principled Generation of Expressive Behavior in an Interactive Exhibit, in Velasquez J.D.(ed.), *Workshop: "Emotion-Based Agent Architectures" (EBAA'99)*, , pages 94-98, Third International Conference on Autonomous Agents (Agents '99), Seattle, WA, USA.
- Petta, P. (2003). The Role of Emotions in a Tractable Architecture for Situated Cognizers. In Trappl, R., Petta, P. and Payr, S. eds. (2003). *Emotions in Humans and Artifacts*, pages 251-287. Cambridge, MA: The MIT Press.
- Pfeifer, R. (1993). Studying emotions: Fungus Eaters. *Proc. First European Conference on Artificial Life (ECAL'93)*, pp. 916-927. ULB, Bruissels, Belgium, May 24-26.
- Pfeifer, R. and Schreier, C. 1999. *Understanding Intelligence*. Cambridge, MA: MIT Press.
- Poggi, I. (2001). Mind markers. In C.Mueller and R.Posner, editors, *The Semantics and Pragmatics of Everyday Gestures*. Berlin Verlag Arno Spitz, Berlin, 2001.
- Prendinger, H., S. Descamps, and M. Ishizuka (2002). Scripting affective communication with life-like characters in web-based interaction systems. *Applied Artificial Intelligence*, 16(7- 8): 519–553.
- Preston, S.D. and de Waal, F.B.M. (2002). Empathy: Its Ultimate and Proximate Bases. *Behavioral and Brain Sciences*, 25: 1-20.
- Picard, R. W. (1997). *Affective Computing*. MIT Press, Cambridge, MA.
- Picard, R.W. (2003). What Does It Mean for a Computer to “Have” Emotions? In Trappl *et al.* (2003), pages 213-235.
- Piwiek, P. (2002). An annotated bibliography of affective natural language generation. Technical report, University of Brighton, July 2002.
- Rao, A.S. and Georgeff, M. (1995). BDI-Agents, from Theory to Practice. In *Proc. 1<sup>st</sup> Intl. Conference on Multi-Agent Systems (ICMAS-95)*, pages 312-319.
- Reeves, B. and Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York, NY: CSLI Publications and Cambridge University Press.
- Reilly W.S., Bates J. (1993). Emotion as part of a Broad Agent Architecture, in Sloman A., Read T. (eds.): *Workshop on architectures underlying motivation and emotion (WAUME'93)*, Univ. Birmingham, 1993.
- Reilly W.S.N. (1996). Believable Social and Emotional Agents, Ph.D.Thesis, School of Computer Science, Carnegie Mellon University, Technical Report CMU-CS-96-138.
- Reisenzein R. (2001). Appraisal processes conceptualized from a schema-theoretic perspective: Contributions to a process analysis of emotions, in Scherer K.R., et al.(eds.), *Appraisal Processes in Emotion: Theory, Methods, Research*, pages 187-201, Oxford University Press, Oxford New York.
- Rist, T. and Schmitt, M. (2002). Applying socio-psychological concepts of cognitive consistency to negotiation dialog scenarios with embodied conversational characters. In R. Aylett and L.

- Cañamero, eds., *Proc. of AISB'02 Symposium on Animated Expressive Characters for Social Interactions*, pages 79–84. Imperial College, London, April 2002.
- Rolls, E.T. (1999). *The Brain and Emotion*. New York: Oxford University Press.
- Rolls, E.T. and Treves, A. (1998). *Neural Networks and Brain Function*. Oxford University Press.
- Roseman I.J., Antoniou A.A., Jose P.E. (1996). Appraisal Determinants of Emotions: Constructing a More Accurate and Comprehensive Theory, *Cognition and Emotion*, 10(3): 241-277.
- Sander, D., Grafman, J., and Zalla, T. (2003). The Human Amygdala: An Evolved System for Relevance Detection. *Reviews in the Neurosciences*, 14: 303-316.
- Scherer, K. (2001). Appraisal Considered as a Process of Multilevel Sequential Checking. In K. Scherer, A. Schorr and T. Johnstone, eds., *Appraisal Processes in Emotion*, pages 92-120. Oxford University Press.
- Slovan, A. (2003). How Many Separately Evolved Emotional Beasts Live within US?. In Trapp, R., Petta, P. and Payr, S. eds. (2003). *Emotions in Humans and Artifacts*, pages 35-114. Cambridge, MA: The MIT Press.
- Smith, C.A. (2004). A Functional Perspective on Emotion Elicitation: Some Considerations for the Development of Emotional Architectures. In Hudlicka and Cañamero (2004), pages 135-143.
- Smith C.A. and Lazarus R.S. (1993). Appraisal Components, Core Relational Themes, and the Emotions, *Cognition & Emotion*, 7(3&4): 233-270.
- Staller A. and Petta P. (1998). Towards a Tractable Appraisal-Based Architecture for Situated Cognizers, in Cañamero D., et al. (eds.), *Grounding Emotions in Adaptive Systems*, Notes of the Workshop at the 5th International Conference of the Society for Adaptive Behaviour (SAB98), pages 56-61, Zurich, Switzerland, August 21.
- Staller A., Petta P. (2001). Introducing Emotions into the Computational Study of Social Norms: A First Evaluation, *Journal of Artificial Societies and Social Simulation*, 4(1).
- Steels, L. (1994): The Artificial Life Roots to Artificial Intelligence. *Artificial Life Journal*, Vol 1,1. MIT Press.
- Stern, A. (2003). Creating Emotional Relationships with Virtual Characters. In R. Trapp, P. Petta and S. Payr, eds. (2003). *Emotions in Humans and Artifacts*, pages 333-362. Cambridge, MA: The MIT Press.
- Tinbergen, N. (1963). On Aims and Methods of Ethology. *Z. Tierpsychol. Beih.* 20: 410-433.
- Trapp R. and Petta P.(eds.) (1997) *Creating Personalities for Synthetic Actors*, Springer, Berlin/Heidelberg/New York/Tokyo, LNAI 1195, 1997.
- Trapp, R., Petta, P. and Payr, S. eds. (2003). *Emotions in Humans and Artifacts*. Cambridge, MA: The MIT Press.
- Velásquez, J.D. (1996). Cathexis: A computational model for the generation of emotions and their influence in the behavior of autonomous agents. Master's thesis, MIT Media Lab, September 1996.
- Velásquez, J.D. (1998). Modeling Emotion-Based Decision-Making. In Cañamero, L. (1998), pages 164-169.
- Wehrle, T. (2001). The grounding problem of modeling emotions in adaptive artifacts. In P. Petta & D. Canamero (Eds.). *Grounding emotions in adaptive systems: Volume I* [Special issue]. *Cybernetics and Systems: An International Journal*, 32 (5), 561-580.
- Wehrle, T. and Scherer, K. (1995). Potential Pitfalls in Computational Modeling of Appraisal Processes: A Reply to Chwelos and Oatley. *Cognition and Emotion*, 9: 599-616.
- Wehrle, T. & Scherer, K. R. (2001). Towards computational modeling of appraisal theories. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.). *Appraisal processes in emotion: Theory, Methods, Research*, pages 350-365, New York: Oxford University Press.

## General bibliography

- Arbib M. A., ed., (2003) *The handbook of brain theory and neural networks* (2nd edition). MIT Press.
- Ben-Ze'ev, A. (2000). *The subtlety of emotions*. Cambridge, Ma.: MIT Press.
- Ben-Ze'ev, A. (forthcoming). Emotions as a general mental mode. In R. Solomon (ed.), *Thinking about feeling: Contemporary philosophers on emotion*. Oxford: Oxford University Press.
- Breazeal C. L. (2002). *Designing Sociable Robots*. Cambridge, MA: The MIT Press.
- Brooks R. A. (2002). *Robot: The future of flesh and machines*. Allen Lane, Penguin Press.
- Cañamero, L.D., ed. (1998). *Emotional and Intelligent: The Tangled Knot of Cognition. Papers from the 1998 AAAI Fall Symposium*. Menlo Park, CA: AAAI Press.
- Cañamero, L. (2001a). Building Emotional Artifacts in Social Worlds: Challenges and Perspectives. In *Emotional and Intelligent II: The Tangled Knot of Social Cognition. Papers from the 2001 AAAI Fall Symposium*, pp. 22-30. Menlo Park, CA: AAAI Press.
- Cañamero, L., ed. (2001b). *Emotional and Intelligent II: The Tangled Knot of Social Cognition. Papers from the 2001 AAAI Fall Symposium*. Menlo Park, CA: AAAI Press.
- Cañamero, L. (2001c). Emotions and Adaptation in Autonomous Agents: A Design Perspective. *Cybernetics and Systems* 32(5): 507-529.
- Cañamero, L.D. (2003). Designing Emotions for Activity Selection in Autonomous Agents. In: R. Trappl, P. Petta, S. Payr (eds.), *Emotions in Humans and Artifacts*. MIT Press, Cambridge, MA.
- Cañamero, L. and Petta, P., eds. (2001). *Grounding Emotions in Adaptive Systems*, Vols. I and II; double special issue of *Cybernetics and Systems: An International Journal*, Vol. 32, Nos. 5 and 6.
- Carberry, S. and Schroeder, L. (2002). Toward Recognizing and Conveying an Attitude of Doubt via Natural Language. *Applied Artificial Intelligence* 16(7/8): 495-517.
- Cavalluzzi, A., Carofiglio, V., and de Rosis, F. (2004). Affective advice giving dialogs. In André, E., Dybkjaer, L., Minker, W., Heisterkamp, P. (Eds.), *Affective Dialogue Systems*. LNCS-LNAI 3068. Berlin Heidelberg: Springer-Verlag.
- Damasio, A. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*. New York: Avon Books.
- Damasio, A. (1999). *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*. London: Vintage.
- Dautenhahn, K., Bond, A., Cañamero, L. and Edmonds, B. (2002). *Socially Intelligent Agents: Creating Relationships wit Computers and Robots*. Kluwer Academic Publishers.
- Davidson R., Scherer K. R., & Goldsmith H. (Eds.) (2003). *Handbook of Affective Sciences*. New York and Oxford: Oxford University Press.
- de Rosis, F., ed., (2002) *User Modeling and Adaptation in Affective Computing*. Special Issue of the journal *User Modeling and User-Adapted Interaction*, Vols. 11-4 and 12-1.
- De Sousa, R. (1987). *The Rationality of Emotion*. Cambridge, MA: MIT Press.
- Evans D. and Cruse P., eds. (2004). *Emotion, Evolution and Rationality*. Oxford University Press.
- Frank R.H. (1988). *Passions within Reason*, W.W.Norton, New York.
- Frijda, N. H. (1986). *The Emotions*. Cambridge University Press, Cambridge, UK.
- Frijda, N. and J. Swagerman, Can computers feel? theory and design of an emotional system. *Cognition & Emotion* 1(3):235-257, 1987.
- Goldie, P. (2000). *The Emotions: A Philosophical Exploration*, Oxford: Clarendon Press.
- Goldie, P. (2004). Emotion, Reason and Virtue. In *Emotion, Evolution and Rationality*. In D. Evans and P. Cruse (eds.), Oxford: Oxford University Press,

- Hudlicka, E. and L. Cañamero, eds. (2004). *Architectures for Modeling Emotions: Cross-Disciplinary Foundations. Papers from the 2004 AAAI Spring Symposium*. Menlo Park, CA: AAAI Press.
- Lazarus, R. S. (1991). *Emotion and Adaptation*. New York: Oxford University Press.
- LeDoux, J. (1996) *The Emotional Brain*. New York: Simon & Schuster.
- LeDoux, J. and Hirst, W., eds. (1986). *Mind and Brain: Dialogues in Cognitive Neuroscience*. New York: Cambridge University Press.
- Lewis, M. & J. Haviland-Jones (Eds.), (2000). *Handbook of Emotions*, Second Edition. New York: Guilford Press.
- Marsella S., Gratch J. (2003). Modeling coping behavior in virtual humans: don't worry, be happy, in Rosenschein J.S. et al., Proceedings of the second international joint conference on Autonomous agents and multiagent systems (AAMAS 2003), 14-18 July 2003, Melbourne, Australia, pages 313-320, ACM Press, New York, NY, USA.
- Mueller, C., Grossman-Hutter, B., Jameson, A., Rummer, R., and Wittig, F. (2001). Recognizing time pressure and cognitive load on the basis of speech: an experimental study. In M Bauer, P J Gmytrasiewicz and J Vassileva (Eds), *Proc. 8th International Conference on User Modeling*, Sonthofen, Germany, July 13 to July 17, 2001.
- Ortony, A., Clore, G., Collins, A. (1988). *The cognitive structure of emotions*. Cambridge University Press, Cambridge, England.
- Paiva, A. ed., (2000). *Affective Interactions: Towards a New Generation of Computer Interfaces*. Berlin-Heidelberg: Springer-Verlag, LNAI.
- Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford University Press.
- Parkinson, B. (1995). *Ideas and realities of emotion*. Routledge, New York.
- Philippot P. & Feldman R. S., eds. (2004). *The Regulation of Emotion*. Lawrence Erlbaum Associates.
- Picard, R. W. (1997). *Affective Computing*. MIT Press, Cambridge, MA.
- Prinz, W. (1997) Perception and Action Planning, *European Journal of Cognitive Psychology*, 9(2): 129—154
- Rolls, E.T. (1999). *The Brain and Emotion*. New York: Oxford University Press.
- Scherer, K.R. and Ekman, P., eds. (1984). *Approaches to Emotion*, Erlbaum, Hillsdale, NJ.
- Scherer, K., Schorr, A. and Johnstone, T., eds. (2001). *Appraisal Processes in Emotion*. New York: Oxford University Press.
- Trappl R. and Petta P.(eds.) (1997) *Creating Personalities for Synthetic Actors*, Springer, Berlin/Heidelberg/New York/Tokyo, LNAI 1195, 1997.
- Trappl, R., Petta, P. and Payr, S. eds. (2003). *Emotions in Humans and Artifacts*. Cambridge, MA: The MIT Press.
- Wehrle, T. (2001). The grounding problem of modeling emotions in adaptive artifacts. In P. Petta & D. Canamero (Eds.). *Grounding emotions in adaptive systems: Volume I [Special issue]. Cybernetics and Systems: An International Journal*, 32 (5), 561-580.
- Wehrle, T. & Scherer, K. R. (2001). Towards computational modeling of appraisal theories. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal theories of emotions: Theories, methods, research* (pp. 350-365). New York: Oxford University Press.
- Wyatt T.D. (2003). *Pheromones and animal behaviour: communication by touch and smell*. Cambridge University press.