Integrating Cognition+Emotion+Autonomy

European Integrated Project IST-027819

6th Framework Programme - Cognitive Systems

Title

State of the Art on Emotion

Integrating Cognition and Emotion for Autonomous Systems

Author

Julita Bermejo Alonso

Reference	ASLab-ICEA-R-2006-001
Release	1.0 Draft
Date	2006-04-15
URL	http://www.aslab.org/documents/ASLab-ICEA-2006-001.pdf
Clearance	Consortium

Partners

University of Skövde, SE
Université Pierre et Marie Curie, FR
Centre National de la Recherche Scientifique, IT
Consiglio Nazionale delle Ricerche, FR
University of Sheffield, UK
University of the West of England, UK
BAE Systems, UK
Cyberbotics Ltd., CH
Hungarian Academy of Sciences, HU
Universidad Politécnica de Madrid, ES

State of the Art on Emotion

ASLab-ICEA-R-2006-001 v 1.0 Draft of 2006-04-15

Abstract

A preliminary literature review on what a concept is, where it resides, how to define concepts and how a concept could be acquired.

Keywords

emotion concept, concept theories, concept in agents, emotional architectures

Acknowledgements

We acknowledge the support of the European Commission Cognitive Systems Unit through grant IST-027819 ICEA.

Revisions

Release	Date	Content	Author
0.0	April 15, 2006	Initial release.	J.Bermejo
0.1	June 2, 2006	Corrections.	R.Sanz

Contents

List of Figures

Chapter 1

Introduction

This report aims at providing an insight into the concept of emotions. First of all, to clarify some constraints encountered during the research of the state of the art should be pointed out. Due to the broad scope of the research on emotions in different domains, the approach has been to cover the key ideas and theories related to each domain, and not so thoroughly only one domain. Secondly, to emphasize that the goal was to fully understand the concept of emotion and related terms from a multidisciplinary viewpoint as required within a more general approach of Cognitive Science. However, the ultimate idea is to gain some knowledge to handle emotions when it comes to artificial machines.

Emotion, emotional state, to be emotional, affect, and other related terms are used as part of our daily life and conversations. Whether we really grasp the underlying ideas about emotions is another issue. It is clear that human beings have emotions. The role of such emotions for human survival, adaptation or cognitive processes is still under scrutiny by the community research. There is no doubt that emotions are needed in human beings, nevertheless it remains to be clarified if artificial systems do need them too. This requirement of emotions in artificial systems, to be fully intelligent and human–like, has lead to a great controversy. Some people are reluctant to consider the role of emotions in such systems. However, if one agrees about their necessity, the way we can exploit their utility in artificial systems remains partially unsolved.

What provides human beings with emotions has been a subject long time of interest for researchers. It is not my purpose to unveil the essential features of emotions. However, I have got the inner feeling of the foundational role that emotions have for understanding the nature of cognition. My pursue is not so much related to human beings but to artificial systems (let it aside what an artificial system is). However, learning from human beings and how they handle emotions seems a sensible idea and a possible trace to follow.

Bearing this in mind, the research on emotions is organized as follows. Chapter 2 introduces the approach from a (Cognitive) Psychology standpoint. Researchers have long paid attention to what an emotion is, how emotions are generated and perceived. Different ideas and theories are explained. Chapter

3 explains how emotions are tackled from a (Cognitive) Neuroscience viewpoint. The approach seems to rely more on the physiology and neural correlates than on the emotion processing.

Next, how emotions are regarded within the Artificial Intelligence (or Computer Science) community, is explained in Chapter 4. This domain has greatly benefited from the research made in the aforementioned domains. However, some questions remain to be answered.

Under each domain, some key terms to build up a taxonomy have been considered. Such taxonomy will provide a useful framework to develop artificial systems with emotions.

Finally, some conclusions gathered from the literature review. Strengths and weaknesses of the different theories and domains are analyzed. Despite the research effort in this report, further work should be carried out to fully cover the topic of emotions and cognition. Such ideas are presented in the last section.

Chapter 2

The (cognitive) psychology approach

2.1 Introduction

The aim of this chapter is to introduce the concepts and theories related to emotion as considered within the domains of psychology and cognitive psychology.

The first step would be to clarify what we understand by psychology within this report purposes. In general, **psychology** is defined as the scientific study of mind and behavior [Friedenberg and Silverman, 2006].

Generally speaking, psychology applies what it is called the Scientific Method. It uses an experiment designated to test an hypothesis. Proving a hypothesis allows to build a theory, in this case, a theory of emotions. In the experiments, independent variables (variable to be manipulated) are used to assess dependent variables (variable to be measured). The experiments could be done by using what it is called the experimental group (subjects who receive the independent variable) and a control group (subjects who do not receive the independent variable).

Additionally, when cognitive processes are taken into account, a new discipline appeared. **Cognitive psychology** is defined as the study of knowledge representation and use in human beings, in other words, how the sensory input is transformed, reduced, stored, recovered, and used [Friedenberg and Silverman, 2006]

Psychology and cognitive psychology have long dealt with the concept of emotion. They have paid attention to the analysis of what could be called basic emotions such as anger or fear, to what could be called moods such as depression or anxiety. Several theories of emotions have been developed and will be further developed within these domains.

Moreover, and more recently, the relationship of emotion and cognitive processes is under scrutiny. How emotions determine or influence learning, attention, decision–making, and so forth are being thoroughly studied by researchers within the cognitive psychology field.

The aim is to gain a deeper understanding on the role of emotions in human beings, and even animals, to tackle problems such as emotional disorders, learning difficulties or improving decision–making.

2.2 On a definition of emotion

.

There is not a well–defined or single definition of emotion. Someone said that there are more definitions of emotions than researchers. Therefore, it is only possible to summarize some definitions available in the literature:

"Emotions are short–term, biologically based patterns of perception, subjective experience, physiology, and action that constitute response to specific physical and social problems posed by the environment" [Niedenthal et al., 2005]

"Emotions are prototypical emotional episodes: a complex set of interrelated subevents concerned with a specific object, accompanied by a subjective feeling that is at least accessible to awareness, even if not always at the forefront of one's consciousness"[Gray et al., 2005]. On the same trend, [Winkielman et al., 2005] defines "emotion as a state characterized by loosely coordinated changes in the following five components: (1) feeling–changes in subjective experience; (2) cognition–changes in attentional and perceptual biases; (3) action–changes in the predisposition for specific responses; (4) expression–changes in facial, vocal, postural appearance; and (5) physiology–changes in the central and peripheral nervous systems".

A further definition is provided by [Scherer, 2005] who states that emotion is "an episode of interrelated, synchronized changes in the states of all or most of five organismic subsystems (cognition, neurophysiological support, motivation, motor expression, subjective feeling) in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism."

" Emotions are states elicited by rewards and punishers, that is, by instrumental reinforcers" [Rolls, 2005b]. To explain it briefly, a reward is an affectively positive stimulus. A punisher is an affectively negative stimulus. For a more detailed description see [Rolls, 2005b].

2.3 Theories of emotions

Several theories have been developed on the concept of emotion, the role of emotions or how the emotional processes take place. A summary of the best known theories within (cognitive) psychology follows:

- James–Lange theory [James, 1884], [Lange, 1885] The theory states that within human beings, as a response to experiences in the world, the autonomic nervous system creates physiological events such as muscular tension, an increase in heart rate, perspiration, and dryness of the mouth. Emotions, then, are feelings which come about as a result of these *bodily changes* (other times called *physiological arousal* or just *arousal* for short).

According to James, emotions feel different from other states of mind because they have these bodily responses that give rise to internal sensations, and different emotions feel different from one another because they are accompanied by different bodily responses and sensations. Fear feels different from anger because it has a different physiological response. The mental aspect of emotion is a subordinate of its physiology, not vice versa: we do not tremble because we are afraid or cry because we feel sad; we are afraid because we tremble and we are sad because we cry. Therefore, only when we pay attention or give an interpretation to the sensation we experience an emotion. If the sensation or physiological arousal is not thought of, we do not experience any emotion.

James and Lange arrived at the theory independently. Lange had a similar view, although he emphasized the role of autonomic feedback such as the heart changes in producing emotions.

The underlying idea could be summarize as the following sequence of events:

```
 EVENT  ---> AROUSAL (BODILY CHANGES) ---> INTERPRETATION --->
EMOTION
```

- Cannon-Bard theory of emotion [Cannon, 1927]

  In this case, arousal (or bodily changes) and emotion take place at the same time. The theory does not pay attention to the role of interpretation or thought of the perceived sensation.

  Following the idea of events showed in the former point, the sequence would be:

```
EVENT  ---> THALAMIC AROUSAL --->  AROUSAL (BODILY
CHANGES) and EMOTION
```

  The theory was developed as an outcome to Cannon's criticisms to the James–Lange theory, based on a physiological and neuroscience approach. Therefore, it could be considered to some extent as a neuroscience theory, where the thalamus is responsible for a thalamic discharge that produces both the emotion and the bodily changes.

- Schacter Two-Factor Theory [Schachter and Singer, 1962]

  They posited that emotion is the cognitive interpretation of a physiological response. In other words, an event causes physiological arousal first. One must then identify a reason for this arousal and then one is able to experience and label the emotion.

  The sequence would be as follows:

```
EVENT  ---> AROUSAL  --->   REASONING ---> EMOTION
```

- Somatic Marker Hypothesis [Damasio, 1994] It is a new view similar to James–Lange theory, which argues that after reinforcers have been evaluated, a bodily response (somatic marker) occurs, leading then to a bodily feeling.

  The purpose of the somatic marker is to drive attention to a negative outcome on someone's actions. The signal may lead to reject the negative course of action, and to choose among other alternatives. The automated signal or somatic marker will remain, helping out in future decision–making by reducing the number of alternatives. Somatic markers include both visceral and nonvisceral sensations.

  As Damasio explains it shortly:

  > "Somatic markers are a special instance of feelings generated from secondary emotions. Those emotions and feelings have been connected, by learning, to predicted future outcomes of certain scenarios"

  .

  Probably it is not a theory in the full sense, but it has been introduced here as many later developments lie on its ideas and assumptions.

- Appraisal theory [Frijda, 1986], [Oatley and Johnson-Laird, 1987], [Ortony et al., 1988], [Lazarus, 1991], [Scherer, 1999]

  Originally, the appraisal theory claims that emotions are elicited and differentiated based on someone's subjective evaluation or *appraisal* of the personal meaning of an object or event according to certain number of criteria. In other words, emotions are outcomes to particular situations.

  The theory has been widely developed by several researchers (see [Scherer, 1999] for a detailed review) with a different approaches when it comes to the appraisal dimensions: criteria, attributions, themes or meanings:

  – Criteria: individuals use a fixed set of criteria to evaluate the significance of the events happening to them. These criteria can be classified as intrinsic characteristics of the objects or events; the significance of the event for the individual's goal; the individual's ability to cope with the consequences of the event, and the compatibility of the event with personal values and rules.

  – Attributions: the emotion–antecedent appraisal is analyzed focusing on the nature of the causal attributions.

  – Themes: different researchers have focused on the possible link between the elicitation of a specific emotion based on the identification of a specific pattern of goal–relatedness of an event.

  – Meanings: researchers have analyzed the semantic nature underlying the use of specific emotion terms.

- Goal–oriented approach [Oatley and Jenkins, 1996]

  The goal–oriented approach considers cognition as a key element to elicit emotions. The underlying idea is that emotions arise from evaluations of events relevant to goals. The emotion is caused by a evaluation or analysis of an event. Each type of evaluation causes a distinct signal that reflects the priority of the goal, which then influences the behavior produced.

- Reinforcement contingency [Rolls, 2005b] Emotions are caused by different reinforcement contingencies by using instrumental reinforcers such as rewards (positive reinforcer) or punishers (negative reinforcer). Any emotion could therefore be described by a combination of reinforcement contingency, intensity of reinforcers, multiple reinforcement associations, different primary reinforcements and different secondary reinforcements, and the available behavioral responses.

2.4 Methodologies to assess emotions

Several methodologies are used in cognitive psychology to assess emotions [of Duke,]:

- To induce the desired state, then to give the subject specific task and to measure the performance.
- Emotional Stroop task to test attention and the interference among multiple processes
- Physiological measures such as galvanic skin response, heart rate, blood samples and facial musculature.
- Imaging techniques such PET scan or MRI.
- Multi–dimensional scaling to classify emotions according to their similarities.

2.5 Cognition and emotion integration

Ortony et al. [Ortony et al., 1988] define the cognition–emotion interaction as follows:

> To say that emotions arise from cognition is to say that they are determined by the structure, content, and organization of knowledge representations and the processes that operate on them.

Some authors consider that emotion and cognition could be integrated as long as they are consider separable, i.e, emotion is not an intrinsic aspect of cognition and vice versa [Gray et al., 2002]. The interaction cognition–emotion has been studied within this domain as a consideration on the effects of different emotions on cognition and vice versa. Different aspects such as attention, memory, and performance have been considered.

A different viewpoint is provided by [Niedenthal et al., 2005], where embodiment is a requirement in the perception and the use of knowledge of emotions. By embodiment, it should be understood the bodily states (e.g. postures, facial expressions, and uses of the voice) that appear when perceiving an emotional stimulus and the later use of emotional information. The purpose is to prove that these bodily states are unconscious whereas the feeling states are conscious. The same trend is followed by [Prinz, 2005].

Further research has addressed the emotion–cognition interactions when it comes to control dilemmas [Gray et al., 2005]. By such, it should be understood how emotion influences the exertion of cognitive control and the direction of selective attention. Cognitive control refers to the "regulation of thought, feeling, and behavior by actively maintained, internal representations of context information, such as a goal that, when held actively in mind, changes the way in which other information is processed". The authors provide some conclusions regarding conflicts in control dilemmas based on their own research as well as others:

- Analytic vs. heuristic processing: emotion as promoting systematic processing (negative moods) or heuristic processing (positive moods).
- Global vs. local attentional focus: attention can either focus on details or on the whole picture. The role of emotion is not fully clear from research, where it is either considered to narrow attention or to bias it toward a global focus.
- Distractibility vs. perseveration: research shows that a presentation of positive images during a task–switching paradigm led to increased switch costs (perseveration) as when affectively neutral images were used, and to decreased switch–costs (distractibility) when the images were affectively positive.
- Speed vs. accuracy: anxiety has been proved to influence speed versus accuracy.
- Emotion on risk taking: emotion influence how risk–reward tradeoffs are considered.
- Self–interest vs. group interest: emotions are strongly social in nature. Research has shown that emotions influence towards cooperation.
- Distinction of emotional phenomena: how emotions interact with cognition depends highly on the type of emotion. Therefore, care should be taken when analyzing the influence of emotion or moods on cognitive processes.

An emotion episode consists of (1) the detection and evaluation of the significance of a stimulus for the individual; (2) the preparation of response tendencies; and (3) the integration of evaluative and proprioperceptive information, resulting in subjective feeling states [Scherer, 2005]. Scherer has proposed a model of the process that underlies the emergence of integrated representations of central processing and proprioceptive feedback into consciousness

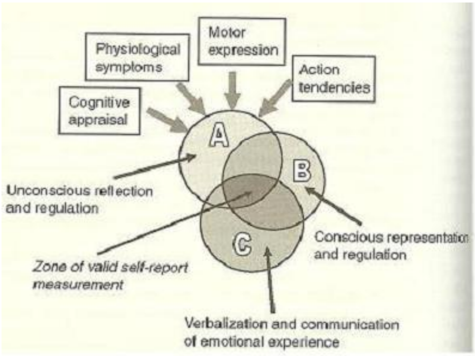

Figure 2.1: Three modes of the representation of changes in emotion components

The first circle (A) could be described as *integrated process representation*. The second circle (B) could be described as *qualia*, i.e., the quality and intensity of the conscious feeling state generated by the eliciting event. The third circle (C) could be labeled as *verbally report* to represent the individual's ability to report with words the subjective experience during the emotion episode.

Some additional findings are as follow [of Duke,]:

Mood congruent recall Cognitions with similar affective tone are retrieved more easily, as memory retrieval is highly influenced by current mood.

Emotion–induced cognitive biases Cognition processes could be influenced by emotions such as anxiety (faster response times, reduced memory consolidation, quick decision making), depression (negative interpretations, perception of less control), and positive affect (influence on cognitive integration, more efficient decision–making, and problem–solving).

Examples of this type of research can be found in [Gray et al., 2002], where functional MRI techniques were used to examine the conjoint effects of emotional states and cognitive tasks on brain activity (focusing on *lateral prefrontal cortex as a possible site for emotion–cognition integration*). Other study has focused on assessing the influence of somatic states related to emotions on a cognitive process such as learning by measuring the skin conductance response [Carter and Pasqualini, 2004].

2.6 A preliminary taxonomy for (cognitive) psychology

- Emotion [Damasio, 1994]: collection of changes in body state connected to particular mental images. See other definitions.

- Emotion type [Ortony et al., 1988]: a distinct kind of emotion that can be realized in a variety of recognizably related forms.

- Types of emotions:

  - Basic (innate, preorganized, primary) emotion: it represents a distinct mode of action tendency. It is physiologically distinguishable [of Duke,].
  - Secondary emotion: experience of feelings (they occur when one begins to form systematic connections between categories of objects and situations as well as primary emotions)
  - Approach emotion: approach motivated emotions such as enthusiam or desire [Gray et al., 2005].
  - Withdrawal emotion: withdrawal related emotions such as fear and anxiety [Gray et al., 2005].

- Bodily changes (physiological arousal, arousal, emotional body state [Damasio, 1994]) [James, 1884]: changes in expression and body response to an event or object

- Feeling of emotions [Damasio, 1994]: experience of changes in body state in juxtaposition to the mental images that initiated such changes.

- Types of feelings [Damasio, 1994]:

  - Feeling of basic universal emotions: associated to basic emotions such as happiness, sadness, anger, fear and disgust.
  - Feeling of subtle universal emotions: associated to subtle variations of basic emotions, such as euphoria and ecstasy (variations of happiness); melancholy and wistfulness (variations of sadness); panic and shyness (variations of fear).
  - Background feeling: the body state prevailing between emotions. A subtype is *mood* which could be considered a background feeling sustained over a period.

- Valence [Charland, 2005], [Winkielman et al., 2005]: the idea that emotions and affects can be classified as positive or negative.

- Eliciting conditions[Ortony et al., 1988]: situational description of the conditions under which the emotion can be trigerred.

- Coping strategy: conscious and selected *mechanism* applied through an act of will for unpleasant or threatening emotions [of Duke,].

- Defense strategy: unconscious *mechanism* for warding unpleasant, threatening, and dangerous emotions. Types are psychotic or neurotic (sublimation, suppression, and humor).

- Techniques (include all the techniques described before).

- Emotion state [Charland, 2005]: what is common to a certain set of evaluative representations, attitudinal behaviors, and physical states.

- Emotion experience [Charland, 2005]: it refers to and includes both the phenomenological aspect of an emotion state and second–order awareness of this experience.

- Eliciting conditions [Ortony et al., 1988]: situational description of the conditions under which an emotion can be triggered.

- Threshold [Ortony et al., 1988]: when eliciting conditions of an emotion are satisfied, there is a context–sensitive emotion–specific threshold. An emotion will only be experienced if its threshold is exceeded.

To clarify the terminological confusion existing within the field of emotion research, Scherer [Scherer, 2005] has proposed the following classification of affects and their attributes:

Design feature **Type of affective state:** brief definition (*examples*)	Intensity	Duration	Synchronization	Event focus	Intrinsic appraisal	Transactional appraisal	Rapidity of change	Behavioral impact	Induction via art/music
Preferences: evaluative judgments of stimuli in the sense of liking or disliking, or preferring or not, over another stimulus (*like, dislike, positive, negative*)	L	M	VL	VH	VH	M	VL	M	H
Utilitarian emotions: relatively brief episodes of synchronized response of all or most organismic subsystems to the evaluation of an external or internal event as being of major significance for personal goals and needs (*angry, sad, joyful, fearful, ashamed, proud, elated, desperate*)	H	L	VH	VH	M	VH	VH	VH	H
Aesthetic emotions: evaluations of auditory or visual stimuli in terms of intrinsic qualities of form or relationship of elements (*moved, awed, surprised, full of wonder, admiration, bliss, ecstasy, fascination, harmony, rapture, solemnity*)	L–M	L	M	H	VH	L	VH	L	VH
Mood: diffuse affect state, most pronounced as change in subjective feeling, of low intensity but relatively long duration, often without apparent cause (*cheerful, gloomy, irritable, listless, depressed, buoyant*)	M	H	L	L	M	L	H	H	M
Interpersonal stances: affective stance taken toward another person in a specific interaction, coloring the interpersonal exchange in that situation (*distant, cold, warm, supportive, contemptuous*)	M	M	L	H	L	L	VH	H	L
Attitudes: relatively enduring, affectively colored beliefs and predispositions toward objects or persons (*loving, hating, valuing, desiring*)	M	H	VL	VL	L	L	L	L	M
Personality traits: emotionally laden, stable personality dispositions and behavior tendencies, typical for a person (*nervous, anxious, reckless, morose, hostile, envious, jealous*)	L	VH	VL	VL	L	VL	VL	L	L

Design feature definitions for major types of affect. VL, very low; L, low; M, medium; H, high; LH, very high. From Scherer (2004b). Copyright 2004 by the *Journal of New Music Research* (www.tandf.co.uk). Reprinted by permission.

Figure 2.2: Definitions for major types of affect

2.7 Conclusions

A first comment when it comes to the concept of emotion within the (cognitive) psychology domain is that of too much information available. Emotion has been a major topic within the domain, therefore much attention has been paid to it.

The literature review has, therefore, served to a personal purpose of a deeper understanding on existing concepts and theories. It does not attempt to provide a full discussion on the adequacy or shortcomings of existing theories. Therefore, the approach of the former sections is to present to the reader with a wide summary of ongoing research on the topic.

An initial conclusion is that of the concept of emotion evolving with time. From initial definitions only considering bodily changes (what later has been defined as embodiment) to later ones, which take into account cognitive processes.

In the former case, the reaction to an external stimulus play a key role. The bodily changes are used both to perceive the emotion and to provide a fundamental way of representing knowledge about emotion (for oneself's sake or for someone else's). Postures, facial expressions and other bodily changes are, therefore, the external representation of emotions.

In the latter case, the stimulus has no intrinsic value. The meaning of the stimulus (which could be an object or event) is determined by a particular organism (either human or animal) in a particular context at a particular point in time. Therefore, how the stimulus is taken into account by the organism and how it influences cognitive processes such as attention or decision–making. In this sense, emotions could be considered as adaptations which allow us to solve problems. Hence, how emotions are classified or distinguished, has evolved from physiology to cognitive aspects.

A further conclusion refers to the terminology and a possible taxonomy. Emotion, affect, moods and other related terms seem to be interchangeable among researchers. If there is not agreement among experts, it seems almost an herculean task to come up with a possible taxonomy within the domain. An attempt has been made to provide such taxonomy. However, further efforts should be made and ontological commitments should be addressed.

What appears to be clear is that research made within the (cognitive) psychology domain has served as starting point for later research, mainly within the Cognitive Science and Artificial Intelligence community. Learning from human beings and animals seem a sensible trend to follow to develop machines with emotions.

Chapter 3

The (cognitive) neuroscience approach

3.1 Introduction

Brain anatomy and function have been long studied. However, advances both in imaging techniques and our understanding of the brain have allowed a further development of the discipline.

Neuroscience is defined as the study of nervous system anatomy and physiology. The later trend to integrate biology and cognition is widely known as **cognitive neuroscience**. Its aim is to explain the structures and physiological processes that underlie certain and specific cognitive functions [Friedenberg and Silverman, 20 Cognitive functions such as recognition, attention, memory, and problem–solving are addressed.

3.2 On a definition of emotion

A possible definition is provided by [Phelps, 2005]:

> Emotion (is used) when referring to the reaction to stimuli that elicit a physiological arousal response, usually due to their aversive nature.

3.3 Theory of emotions

Cognitive scientists and neuroscientists have not always paid enough attention to emotions [Damasio, 1999]. Only brain and mind have been considered from a mere physiological viewpoint, and moreover, only recently.

3.3.1 Damasio's theory of emotions

Damasio [Damasio, 1999] distinguishes three different stages along a continuum:

- A state of emotion: which can be triggered and executed nonconsciously
- A state of feeling: which can be represented nonconsciously
- A state of feeling made conscious: known to the organism having both emotion and feeling

This allows for a deeper discussion and investigation to characterize such phenomena. The distinction between the different stages provide room for the existence of emotions in nonhuman beings, which experiment emotions but find it difficult, if not impossible, to be aware of their feelings.

Moreover, he establishes a clear distinction among feeling and emotion. Whereas the first is an inner experience, the latter is an external one. Therefore, one cannot observe a feeling unless perceiving one's own feelings. Only emotions and their external outcome could be observed.

Emotions cannot easily be controlled. We can partially control the expression of an emotion. Moreover, he considers that animals or humans are innately wired to respond with an emotion when certain stimuli such as size, motion or sounds are perceived on their own or combined [Damasio, 1994].

Damasio considers three different type of emotions [Damasio, 1994], [Damasio, 1999]:

- Primary or universal emotions: the first idea we have on what an emotion is. They include happiness, sadness, fear, anger, surprise, or disgust. The amygdala is believed to be the key player in this type of emotions [Damasio, 1994].

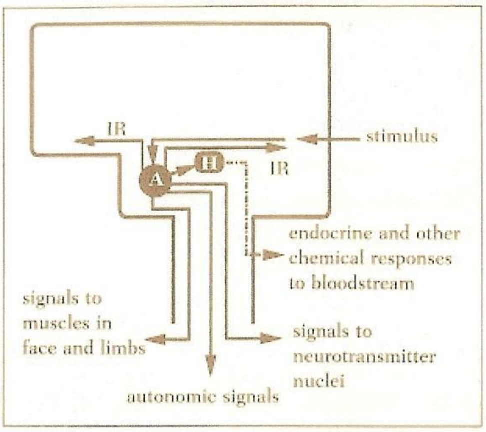

Figure 3.1: Primary emotions

The black perimeter in the figure represents the brain and brain stem. A stands for the amygdala. IR stands for the internal response. H stands for the hypothalamus.

- Secondary or social emotions: a further notion to deal with our role as social beings. They include embarrassment, guilt, or pride.

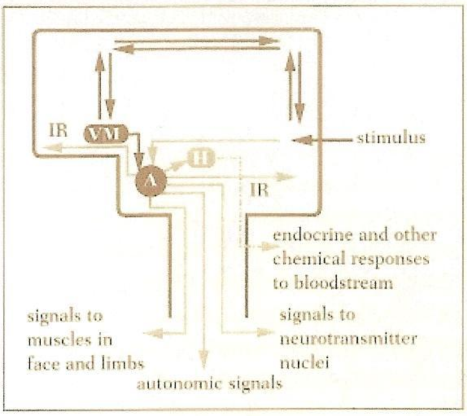

Figure 3.2: Secondary emotions

The black perimeter in the figure represents the brain and brain stem. A stands for the amygdala. IR stands for the internal response. H stands for the hypothalamus. VM stands for the frontal cortex.

- Background emotions: a third type introduced to allow a description of internal states triggered either by certain conditions of internal state or interaction with the environment or a combination of both. They include well-being, dread, relaxation, fatigue.

Damasio states some features of emotions and feelings:

- Emotions are collections of chemical and neural responses.
- Background and culture influence the meaning and expression of emotions.
- The neural correlate for emotion is a restricted ensemble of subcortical regions, from the brain stem to the higher brain.
- Emotions have stereotypicity, automaticity and regulatory purposes.
- All emotions influence changes on body and brain. The collection of this changes constitutes the substrate for the neural patterns which become feelings of emotions.

To summarize the previous ideas, Damasio composes the primary and secondary emotions, as well as feeling with bodily responses as shown in Fig. 3.3.

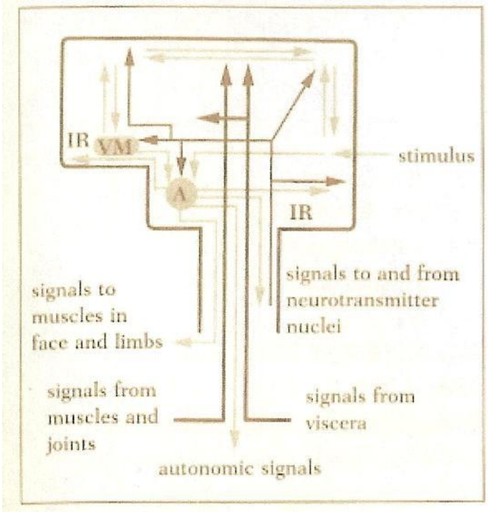

Figure 3.3: Emotion and feeling

The biological function of emotions

Damasio [Damasio, 1999] considers that emotion are part of the bioregulatory devices as part of our survival equipment.

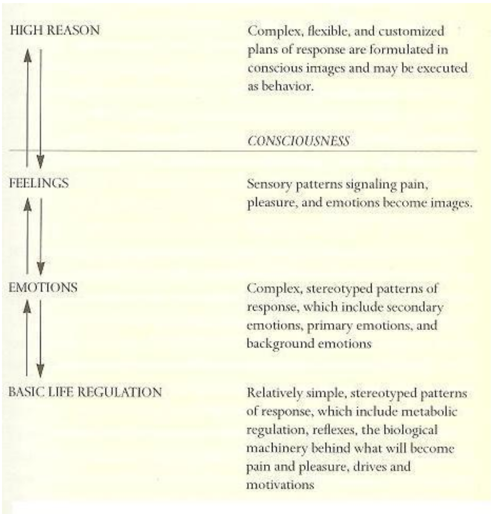

Figure 3.4: Levels of life regulation

The biological function of emotions is twofold. On the one hand, the production of a specific reaction to inducing situation (e.g. run, become immobile both in animals and humans). On the other hand, the regulation of the internal state of the organism to allow it to be prepared for the specific reaction (e.g. increased blood flow to legs to allow running; changing heart and breathing rhythms in case of standing still).

3.3.2 LeDoux and Phelps works

LeDoux, Phelps and colleagues [LeDoux, 2000], [LeDoux, 2002], [Phelps, 2005], [Phelps and LeDoux, 2005], [Fellous and LeDoux, 2005] have long worked on research on cognition and emotion from a neuroscience viewpoint. His research has addressed the neural correlates for emotions. It was thought that the limbic system was the way the brain makes emotions.

LeDoux's works mainly on the emotion of fear have allowed to define brain circuits responsible for such emotion. The approach is that of *fear conditioning*: (A) fear conditioning involves presenting an unconditioned stimulus at the end of a neutral conditioned stimulus; (B) After conditioning, the conditioned stimulus elicits a wide range of responses (behavioral and physiological) similar to those that take place when an animal encounters a fear stimulus.

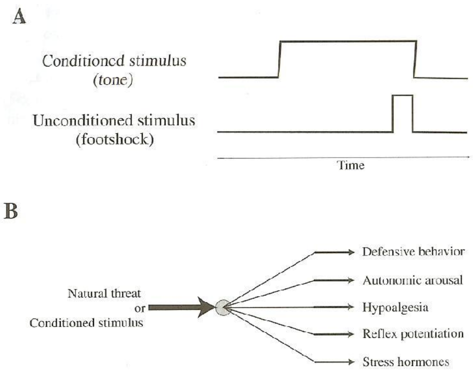

Figure 3.5: Fear conditioning

His research has proven the fundamental role of the amygdala when it comes to emotions (or more particularly for the emotion of fear). It remains to address in detail how fear processing by the amygdala is influenced or can influence other cognitive processes such as perception, attention, and memory functions of the cortex [Fellous and LeDoux, 2005]. In other words, the interaction between cognition and emotion with the amygdala playing a major role.

In the figure, it is shown how the amygdala (A) receives inputs only from the late stages of cortical sensory processing (thick arrow). Once the amygdala is activated, it regulates the cortical areas. The amygdala also activates arousal systems, which can then influence sensory processing. The amygdala also interacts with the medial prefrontal cortex (mPFC), which joint to the dorsolateral prefrontal cortex (dlPFC) influence cognition and behavior, sending connections to the amygdala to regulate it and its fear reactions.

He has also addressed the integration of cognition, emotion, and motivation in the so called *mental trilogy* [LeDoux, 2002].

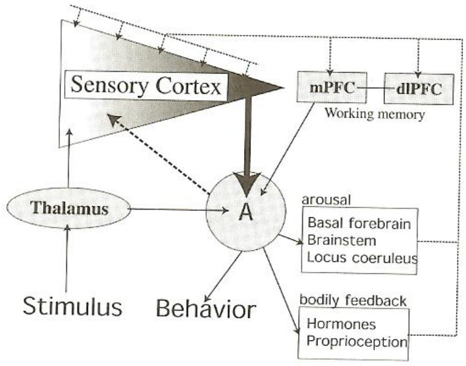

Figure 3.6: The role of the amygdala in emotion

3.3.3 Rolls research

Following his research on a theory of emotions based on rewards and punishers (see previous Section), Rolls has addressed the response to rewarding and punishing stimuli when it comes to the brain [Rolls, 2005b], [Rolls, 2005a].

He has shown that there are two routes to action performed in relation to reward or punishment in humans:

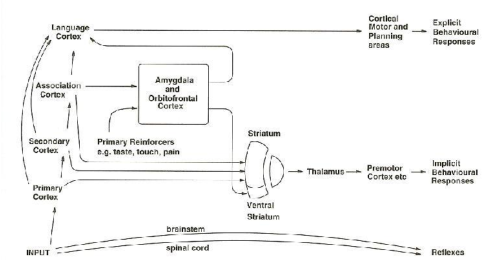

Figure 3.7: Dual routes of action

- The first route is via the amygdala and the orbitofrontal cortex. The stimulus is evaluated to establish its reward according to different factors. Once this is done, the behavior response is initiated as an approach towards or withdrawal the stimulus.

- The second route is via the language systems of the brain. It could be considered similar to processing several if..then statements, to implement a plan to obtain a reward. It is exclusive to humans, who can process syntax rules and symbols. Probably, this second route is related to consciousness.

3.4 Cognition and emotion interaction

The interaction among emotion and cognition within the neuroscience domain has been analyzed from a neural systems viewpoint. With such aim, studies and research on patients suffering from different brain damage as well as different techniques (e.g. instructed fear, reappraisal instruction, attentional blink paradigm, etc) have been carried out.

In this sense, research made considers the role of the human amygdala as a specialized structure for emotions as well as its interaction with processes of cognition and awareness [Phelps, 2005]. On the one side, the amygdala can be influenced by symbolic representation, cognitive control, and conscious interpretation. On the other side, emotion, through the amygdala, can influence cognitive mechanisms and conscious awareness of both events and stimuli. More precisely, the amygdala has been proved to modulate long–term retention of memory and to influence attention and perception. The main conclusion is that of the neural systems of emotion and cognition being both independent and interdependent [Fellous and LeDoux, 2005].

3.5 Conclusions

As several authors have recently pointed out [Damasio, 2004], emotion has traditionally been ignored in cognitive science because it was, and still commonly is, associated with irrationality, and consequently few researchers have considered it worthwhile to integrate emotions in cognitive systems architectures. However, recent findings indicate quite the opposite, i.e. that emotion in fact plays a crucial role in rational decision-making [LeDoux, 2000]. For example, studies of human subjects with specific brain lesions that significantly reduce the capacity to experience emotions show that the subjects' capacity to make rational decisions was severely impaired although intellectual capacities were (seemingly) left intact [Damasio, 1994], [Damasio, 1999].

Emotion can be viewed as a flexible adaptation mechanism that has evolved from more rigid adaptational systems, such as reflexes and physiological drives . The flexibility of emotion is obtained by decoupling the behavioral reaction from the stimulus event. The heart of the emotion process thus is not a reflex like stimulus-response pattern, but rather the appraisal of an event with respect to is adaptational significance for the individual, followed by the generation of an action tendency aimed at changing the relationship between the individual and the environment.

However, it should be noted that the above view of emotions and their integration in cognitive systems architectures is still far from widespread in cognitive systems research.

The connection between emotion and cognition is particularly clear in what Damasio calls feeling, that is

> the mental representation of the physiologic changes that occur

during an emotion.

That means, while emotions involve bodily reactions, feelings (mental representations/images of those reactions) allow the individual to temporarily detach its cognitive processes from its immediate bodily reactions, as, for example, in the anticipation of such reactions or the planning of behavior. Damasio refers to the underlying cognitive process as an ”as if body loop”, i.e. a neural ”internal simulation” that uses the brain’s body maps, but bypasses the actual body. He elaborates the relation between emotions and feelings as follows:

> The essence of feelings of emotion is the mapping of the emotional state in the appropriate body-sensing regions of the brain. But feeling an emotion also includes the mapping of changes that occur in the cognitive processing style, as well as the evocation of thoughts that are congruent with the feeling state.

Whereas emotions provide an immediate reaction to certain challenges and opportunities faced by the organism, feeling the emotions provides the organism with a mental alert for the significance of the object that caused the emotion and for the thoughts consequent to responding emotionally [Damasio, 2004]. The adaptive value of feelings comes from amplifying the mental impact of a given situation and increasing the probabilities that comparable situations can be anticipated and planned for in the future so as to avert risks and take advantage of opportunities.

It should be noted that due to the still relative recent interest in emotion from different disciplines, much of the terminology is still not well–defined and the use of central terms, such as emotions and feelings, still varies between researchers (not only within the neuroscience or cognitive science domain, but also in cognitive psychology and the more recent artificial intelligence field).

Chapter 4

The artificial intelligence approach

4.1 Introduction

To define what Artificial Intelligence, a new discipline if compared to the previous ones, is not a simple task. It could be defined as the attempts to build devices to mimic functions, and more recently, thought processes as compared to animals and humans. When related to cognitive science, the aim is to code knowledge and meta–knowledge (facts about knowledge, such as rules to use our knowledge) to assemble systems which will allow to explain intelligence, consciousness, and related cognition processes [Friedenberg and Silverman, 2006].

Artificial Intelligence (AI) and its related field of Computer Science (CS) have lately paid attention to the role of emotions for machines and robots. As expressed by Minsky in The Society of Mind (a usual quote among AI researchers):

> The question is not whether intelligence machines can have any emotions, but whether machines can be intelligent without any emotions.

Although it might look like a tongue–twister, it comprises a controversial idea among AI researchers on emotions. It clearly states the necessity of emotions when it comes to what could be called "artificial" emotions in intelligent machines. Therefore, a great extent of research has been made within the AI community to address this fundamental role of emotions in machines.

Some of the ideas and research made is introduced in the following sections. Firstly, some definitions of emotions are provided.

4.2 On a definition of emotion

Sometimes but not always, emotions within the AI domain are referred to "artificial" emotions. It is unclear whether it is just a terminology inherited from the broader domain of "artificial" intelligence or it comprises deeper ideas. The discussion about a proper terminology is out of the scope of this report. Throughout the chapter, the term emotion will only be used. However, it is worthwhile pinpointing that the adjective artificial goes along with it.

Despite of naming it or not as artificial, there are a myriad of definitions on the concept of emotion. As it happened within other domains, researchers seem not to be able to agree on a sole definition. Therefore, some available definitions are summarized here:

"Emotions are a sequential process comprising the appraisal of the agent global state, the generation of an emotion–signal, and an emotion–response" [Botelho, 2001]. What it is understood by an emotion–signal and an emotion–response will be fully explained in the next section.

4.3 The role of emotions

Emotions have become a major research interest among the AI community. Emotions are considered to be evolutionary engineering solutions required for human–like intelligence. Emotional mechanisms are considered to provide agents and architectures with greater adaptability and robustness, aspects especially important in complex real–time systems.

Different roles have been assigned to emotions (or emotional mechanisms) [Gadanho and Hallam, 2001], [Oliveira and Sarmento, 2003], [Botelho, 2001], [Hudlicka, 2004], [Cañamero and Gaussier, 2005]:

1. *Measurement of degree of success*: agents are provided with emotions as evaluation mechanisms of their performance in the surrounding environment.

2. *Information collectors*: emotions allow to filter relevant data, condense information and guide decision–making. Emotions can also improve the way information is created, conveyed, and understood.

3. *Management mechanism*: emotions influence and allow to manage cognitive capabilities and processes.

4. *Believability of agents*: emotions are viewed as a way to improve it in the sense of agent to appear more life–like to humans. Closely related to expressiveness and communication.

5. *Expressive behavior*: emotions are used to improve interface agents and avatars (focusing on basic emotions by Ekman).

6. *Signals and communication*: Emotions are also used as communication mechanisms that allow a robot to report its internal state to others. Emotions are used to assess situation from others' emotions. They aid to construct the image of self and other.

7. *Adaptive mechanisms*: emotions are used to deal with important events related to survival or to control of interactions with (physical and social) environment.

4.4 Issues in emotion modeling

To specify or model emotions in AI systems (either architectures or agents) is a problem that requires a deep analysis. An initial view on the issues to be considered when modeling emotions was described in [Pfeifer, 1987]:

- Emotion as process: which processes are involved
- Emotion generation: which conditions determine and trigger emotions
- Influence of emotion: how the generated emotion influences the further behavior of the system
- The goal–oriented nature of emotions: the computer models need a explicit specification of goal–oriented structures
- Interaction between subsystems: how different subsystems should be synchronized to handle emotional and non–emotional states
- Emotions as heuristics: emotions help in predictions and decision–making
- Representation of emotions: the key issue of how to represent the emotion in a computer–based system

It could be concluded from the aforementioned points, that modeling emotions implies several elements to consider. With time, some of the aforementioned issues have been clarified. Others, remain under scrutiny.

Further research has established the following elements to consider when modeling emotions are provided in [Hudlicka, 2004]: (1) the context that requires emotion models; (2) methods available to model emotions within cognitive architectures and (3) alternative modeling.

A more detailed analysis is provided in [Cañamero, 2005] where the following issues should be considered when developing emotion–based architectures (and agents):

1. Regarding models: there are different models and theories when it comes to emotions. Do they address the same phenomena? Is the underlying concept/definition of emotion the same in all of them? Is it possible a **general definition of emotions** for modeling?
2. Regarding emotion machinery: computational mechanisms and models are used for AI emotions. Which mechanisms should be used? How different mechanisms can be integrated? How to model the cognition–emotion interaction?

3. Regarding application: the research is done within the AI (agent–based) community. Therefore, which emotions are more suitable for autonomous agents and robots? Which model would be better implemented in an application?

4. Regarding the influence of emotion in cognition: it is somehow clear that there is an emotion–cognition interaction. The remaining point is to assess and quantify how the robot or agent behavior is a consequence of the emotion state. Could it be a result of the robot or agent interaction with the environment?

The aforementioned aspects do not attempt to answer all the questions but to bring to our attention remaining issues for the artificial emotion research.

Additionally, some authors are not only concerned with the modeling itself [Sloman, 2001], [Sloman, 2004]. It is claimed the necessity of a deeper analysis (to overcome the so–called shallow models) and framework (sometimes referred as an ontology) to develop and model emotions within the agent community.

With such purpose, an initial framework for AI emotion research is provided in the following section. Later on, an initial (and possibly incomplete) taxonomy is provided.

4.5 A framework for AI emotions research

The research methodology used within the Cognitive Science and Artificial Intelligence community is the *Computational Modeling approach*. Exploring and modeling what emotions are, it is made from different viewpoints [of Duke,], [Cañamero, 2001], [Cañamero, 2005] which allow to establish a framework to categorize such computational models:

- Inspiration source: It can be distinguished three different perspectives depending on the inspiration source used:

  Semantic–based models They consider the way people use emotional related terms in natural language, to look for a classification and eliciting conditions. Not so widely used in embodied AI.

  Phenomenon–based models They assume the possibility of identifying an emotional state to identify its related phenomena. They have been probed to be useful when emotions are considered to be behavior–producing mechanisms related to particular goals or tasks.

  Design–based models They analyze the way a system should be design to fulfil a particular behavior. They pay special attention to the relationship among the underlying mechanisms, the resulting behavior and the environment where the behavior takes place (to assess the suitability of the model). Models could follow a bottom–up, middle–out or top–down approach.

- Level of abstraction: The models are used either to address either a **single phenomena** or to model **(cognitive) agent architectures)**. The models within the latter category have been developed with different level of abstraction such as *architecture–level models* (they embody emotional processing); *task–level models* (they address a single task related to emotions), and *mechanism–level models* (they attempt to emulate some specific aspect of emotions processing).

- Goal of the research: The goal of the research could follow either an "engineering" or a "scientific" motivation. In the former, **output or black–box models** are developed to build more robust agent–based systems capable of a more efficient and better adaptation to the environment. In the latter, **process models** which address the relationship among emotion and cognition to assess the role of emotions for adaptive behavior.

- Perspective on emotion: Two different kinds of models could be distinguished. A first one, **component–based models** which consider that an artificial system or agent has emotions when it owns a certain number of components that characterize a human (or animal) emotional system. A second one, **functional–models** which focus on the properties of humans (and animals) as well as their environment, which can applied to artificial systems (agent and environment) to own the same functionality.

- Underlying assumptions: The existing research lies on two different underlying assumptions when it comes to the relationship among emotion processing and cognition. On the one hand, emotions are considered to be *secondary* to cognition or even an outcome of cognitive processes. On the other hand, emotions stands *on its own* alongside with cognition processes such as memory, attention, learning and so forth.

- Implementation of the model: A key point to assess any theory of emotion within this domain, it is to consider to what extent the computational model has been developed.

- "Depth" of emotion modeling: Some authors [Sloman, 2001], [Hudlicka, 2004] distinguish between **deep emotion models** which explore architectural features and configurations capable of producing emotions vs. **shallow emotion models** which only reason on emotions.

Most of the research to be described in the following section could be associated within one or more concepts from the aforementioned framework.

4.6 Research on emotions: some examples

- Institute for Systems and Robotics (IRS) at Lisbon: the research group worked on the DARE (Development of Emotion–based Robotic Agents) project between 1999 and 2004 [Gadanho and Hallam, 1998], [Gadanho and Hallam, 200 [Custódio et al., 1999]. The aim of the project was the study and development of methodologies and tools to implement emotional robotics agents capable of dealing with unstructured and dynamic environments. Therefore, the goal was not the optimization of some particular ability, but on general competence to learn, adapt and survive. To test these ideas, a small autonomous robot was used based on technology already developed and tested.

The research followed different trends following the works by Damasio [Damasio, 1994]. In the first one, reinforcement learning was initially studied by Gadanho et al. [Gadanho and Hallam, 1998], [Gadanho and Hallam, 2001], where an autonomous robot needs to adapt to its environment. Emotions are used to determine state transition in a reinforcement–learning system. As outcome, the ALEC (Asynchronous Learning by Emotion and Cognition) architecture was developed [Gadanho and Custódio, 2002]. An initial and simple emotion–based architecture consisted of a goal system and an adaptive system. ALEC architecture added a cognitive system, which provided an alternative decision–making process to the emotion system.

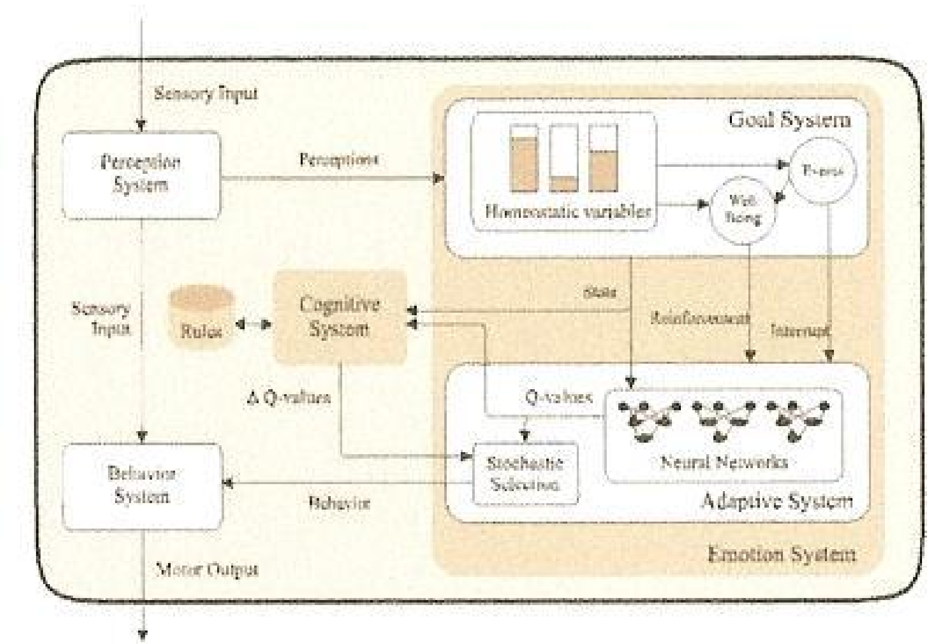

Figure 4.1: ALEC architecture

A second line of research, developed the DARE model [Custódio et al., 1999], [Ventura and Pinto-Ferreira, 1999], concerned with the dual evaluation of the perceptual stimulus. The *stimulus* is evaluated from both a perceptual and a cognitive viewpoint. The former provides a *perceptual image* (essential features meaningful to the agent). The latter provides a *cognitive image* (to allow reconstruction of the original stimulus). Therefore, the model consists of a cognitive and a perceptual layers, which receive the stimulus to be processed in parallel. A Desirability Vector (DV) (each one of its components represent a basic kind of assessment of a stimulus) is also used by the perceptual layer. Different implementations were defined and implemented to test the hypothesis for control and supervision purposes.

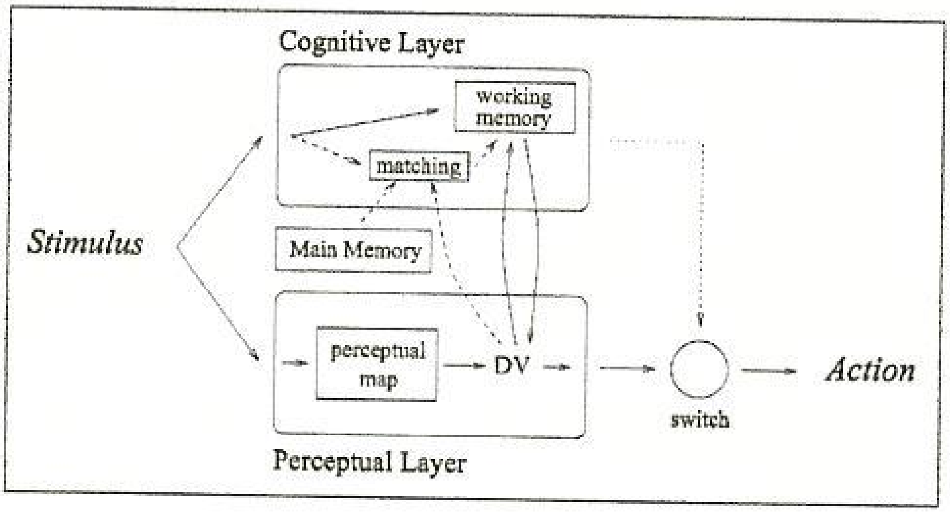

Figure 4.2: DARE model

- Salt and Pepper project [Botelho, 2001]: the aim is to develop and evaluate computer–agent architectures for general intelligence. To do so, the role of emotion is considered as inspired by the cognitive science and neuroscience, where emotions are used in problem–solving and decision–making skills.

  The emotions considered within the project are not so much human emotions themselves, but as *performance evaluators* (emotion–signal as evaluation of performance in a task) and *attention–shift warnings* (emotion–signal as warning to focus attention to something else).

  An architecture for autonomous agents has been developed within the project. It shows the important role of emotions and their impact on cognitive processes such as the allocation of cognitive resources, attention control and adaptive learning.

- Adaptive Systems Research Group in Hertfordshire (UK): it is worth mentioning it, the research carried out by Dolores Cañamero. Her works could be framed within what is called affective computing (term coined by Picard), with a focus on the human–machine interaction on an emotional basis [Cañamero and Fredslund, 2000], [Cañamero and Gaussier, 2005]. Issues addressed are how to convey intentionality to machine emotions, how to elicit and control machine emotions, as well as how to improve human–machine communication by using emotions.

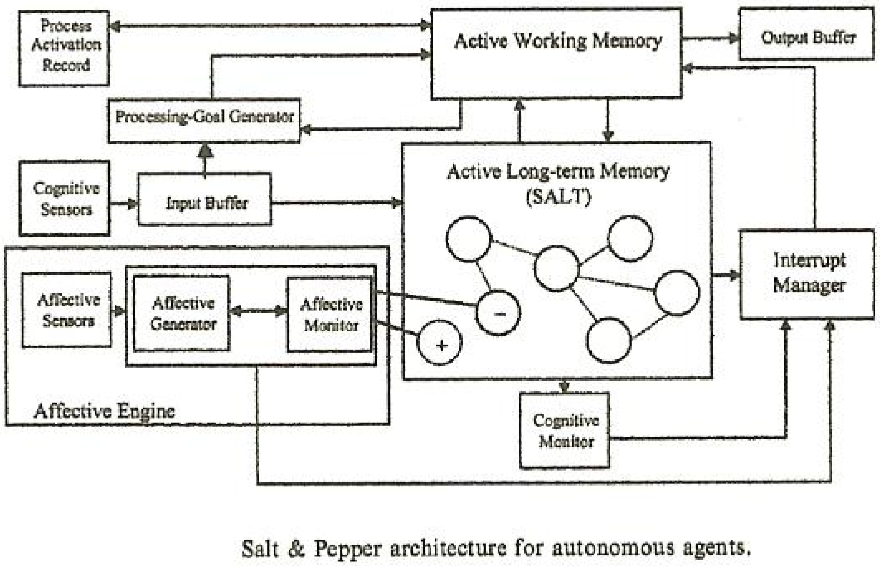

Figure 4.3: Salt and Pepper architecture

Following the same trend, further aspects addressed by Cañamero focus on how to model emotions when it comes to autonomous robots. A review on concepts, existing research and open issues is described in [Cañamero, 2001] and [Cañamero, 2005].

A former research focus on emotions as adaptive mechanism which serve a particular purpose [Cañamero, 1997], [Cañamero, 2001].

- Pyrosim Environment [Oliveira and Sarmento, 2003], [Sarmento, 2004]: the emotion–cognition interaction is studied by means of eliciting or evaluating goals, agents' capabilities and internal states and the environment condition or state. Their emotion–based architecture used new introduced terms such as Emotional Evaluation Functions (a function whose inputs are the information of the environment and the agent internal state. Its output reflects the chances of the agent to achieve a particular goal), Emotional Accumulators (a time dependent process that incrementally stores a percentage of output value of one Emotional Evalutation Function) and a Emotional Structure to connect the former two elements. The focus is on the interaction between a highly cognitive process such as *planning* and emotions. The authors used a simulated environment to test the agent–based architecture (the Pyrosim Environment).

- Cognition and Affect Project, University of Birmigham, UK [Sloman, 1982], [Sloman, 2001], [Sloman, 2004], [Sloman, 2005]: Sloman and his research group have been long working on the topic of emotions when it comes to artificial machines. Their approach is to investigate cognitive mechanisms not in an isolated way, but within cognitive architectures. They

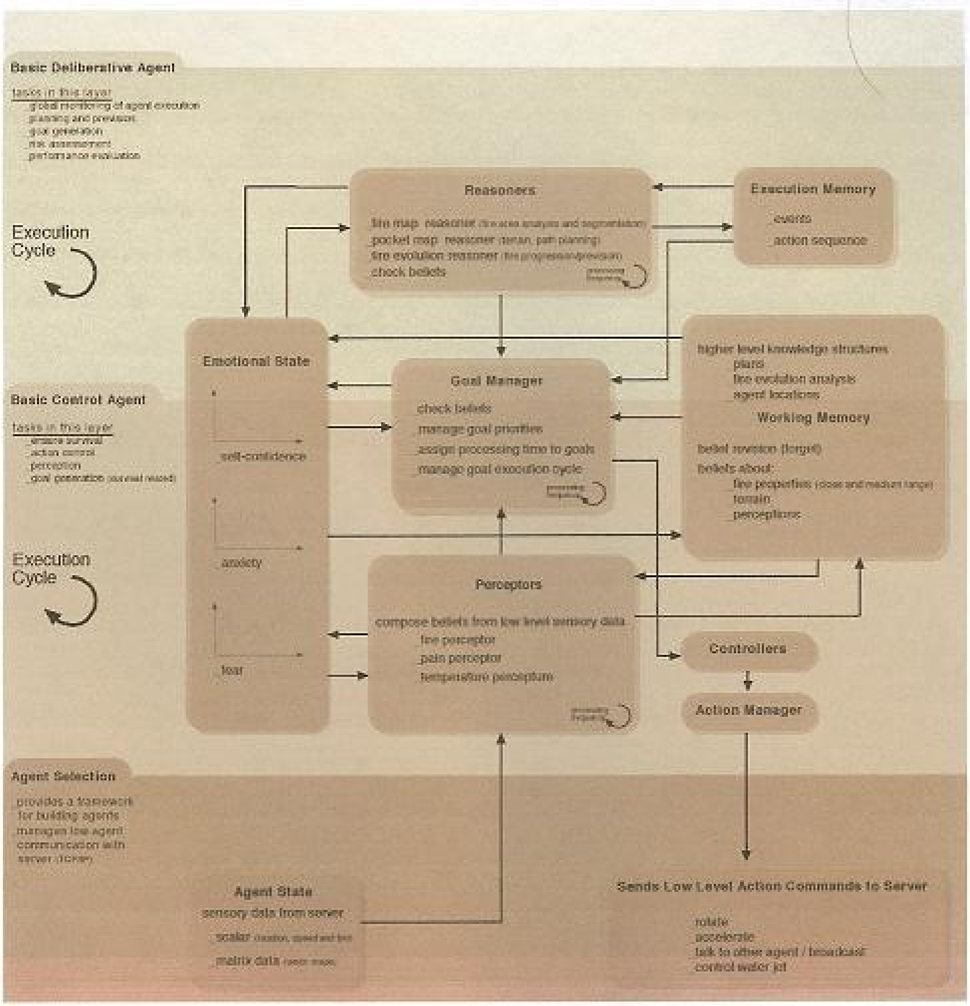

Figure 4.4: Emotion based agent architecture

have explored different architectures(CogAff and H–CogAff) in what they defined as a *design–based approach*, by analyzing requirements and implications of all possible architectures.

As an example, he has defined three different types of emotions: two of them corresponding to the primary and secondary emotions by Damasio; a third one introduced to describe emotions related with thought and attention control (he argues that such emotions are probably exclusive to humans). These three types fulfil the three–layered architecture defined in his research (inspired by Evolutionary theories): Reactive layer, Deliberative layer and MetaManagement layer. Therefore, emotions occur as a result of architectural requirements.

Additionally, Sloman has devoted much time to discuss on emotions from a more "philosophical" viewpoint. He has criticized existing theories of emotions, by arguing that further ontologies should be developed to produce a shared terminology for research. He has also worked on the argument about the necessity of emotions for a machine

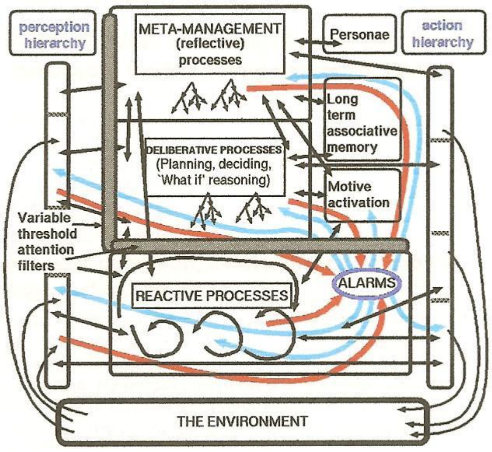

Figure 4.5: CogAff architecture

to be intelligent (as well as trying to clarify Damasio's ideas for artificial systems).

- Velásquez's research [Velásquez, 1997], [Velásquez, 1998]: he developed an architecture, Cathexis, as part of his MSc research. The architecture is modular, flexible and extendable that supports Primary and Secondary emotions (as defined by Damasio). He also developed a robot whose decision–making process is based on the proposed architecture, being a hardware implementation of emotional mechanisms.

  Velásquez consider emotion as biological phenomena which possess a strong relationship to survival and adaptation. He includes both cognitive and non–cognitive components of emotion.

- MAMID Cognitive Architecture [Hudlicka, 2004]: Hudlicka is concerned with the necessity of emotions in cognition. She argues that to model emotion in a particular agent, it should be considered: (1) the context that requires emotion models; (2) methods available to model emotions within cognitive architectures and (3) alternative modeling.

  She has proposed a modeling approach, MAMID, to be used as a methodology to model emotions within the MAMID Cognitive Architecture.

  Such methodology maps a particular state (such attention speed, anxiety, fear, etc) onto specific architecture parameters. These parameters are used by the modules of the architecture (which is designed in a sequential way).

4.7 A preliminary and incomplete taxonomy

- Artificial emotion [Botelho, 2001]: sequential process consisting of the agent's global state appraisal, generation of emotion–signals to regulate the agent's behavior, and an emotion–response.

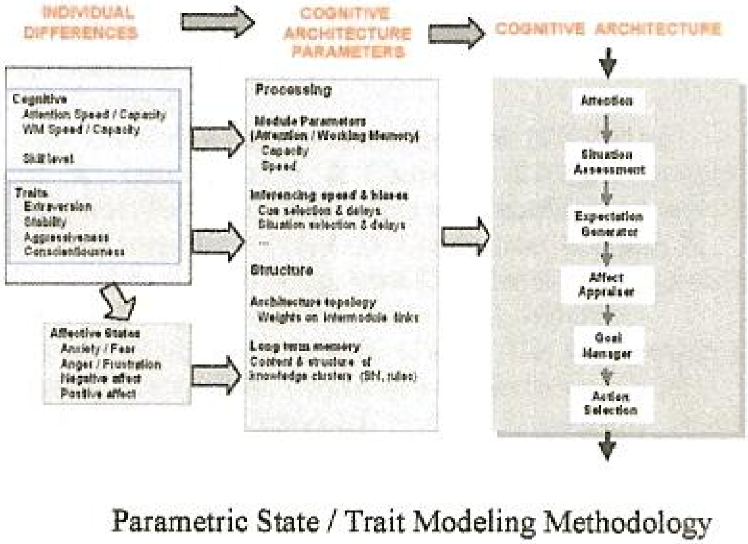

Figure 4.6: MAMID modeling approach

- Emotion–signal[Botelho, 2001]
- Emotion–response [Botelho, 2001]
- Performance evaluator [Botelho, 2001]: an emotion–signal that represents an evaluation of the performance of the agent in a given task.
- Attention–shift warning [Botelho, 2001]: a signal to warn the agent to pay attention to something else.
- Eliciting conditions [Botelho, 2001] (cf. [Ortony et al., 1988]: situational description of the conditions under which the emotion can be trigerred.
- Emotion experience [Botelho, 2001]: combination of emotion signals and the cognitive and behavioral agent outcome.
- Emotional phenomena [Oliveira and Sarmento, 2003], also used by Velásquez.
- Attributes of emotions [Oliveira and Sarmento, 2003]
  - Object/antecedent: pre–condition that trigger the emotional phenomena
  - Intensity: how strong the Agent is influenced by the emotion
  - Duration: time span from the rise to the fall of the emotion
  - Consciousness: whether the agent is conscious or not about the occurrence of the emotional phenomena

- Types of emotions [Oliveira and Sarmento, 2003], [Velásquez, 1998]
  - Specific or primary emotions (similar to Damasio's basic emotions):
    * Examples or instances: anger, fear, joy, surprise, disgust.
    * Object/antecedent: well defined
    * Duration: reduced time span
    * Intensity: strong
    * Consciousness: clear but not immediate
  - Moods or secondary emotions(similar to Damasio's secondary emotions)
    * Examples or instances: anxiety, relaxation, self–confidence, frustration.
    * Object/antecedent: no well defined
    * Duration: from few hours to few days
    * Intensity: no very intense
    * Consciousness: unconscious to agents
  - Emotional dispositions (similar to Damasio's mood, i.e., long–term feeling of emotions). Also called Emotional Phenomena by Velásquez.
    * Examples or instances: chronic anxiety, depression.
    * Object/antecedent: genetics, medications, neurosurgery.
    * Duration: up to several months (but inactive most of the time).
    * Intensity: very intense
    * Consciousness: not defined
  - Other classification [Sloman, 2001]:
    * Primary emotion: outcomes of the interaction between Alarm mechanisms and subsystems of the Reactive and Deliberative layers.
    * Secondary emotion: outcome of Alarm mechanisms evaluating internal cognitive responses
    * Tertiary emotion: outcome of the MetaManagement Layer, related to thought and attention control.
- Roles of emotions [Oliveira and Sarmento, 2003]
  - Measurement of success
  - Information collector
  - Management mechanism
- Emotional elicitation [Oliveira and Sarmento, 2003]: process of evaluation that involves the chances of achieving a goal, the state of the environment, and the internal state and capability of the agent.
- Goal [Oliveira and Sarmento, 2003]: future states to be achieved by the agent as well as implicit states to be maintained.

- Coping potential [Oliveira and Sarmento, 2003]: Capability of the agent to cope with the current state of the environment for achieving one or more of its goals.
- Emotion valence [Gadanho and Hallam, 2001]: an emotion can provide a positive or negative value (as in appraisal theory).
- Signal valence [Botelho, 2001] : a signal can be either positive (if agent global state is favorable to agent motives) or negative otherwise.
- Emotional mechanism [Velásquez, 1998], [Cañamero, 2000]: it consists of a triggering event, an intensity level, an activation threshold, a list of synthetic hormones to be released when activated and a list of physiological manifestations (as described by Cañamero).

4.8 Conclusions

A first conclusion can be drawn when it comes to emotions in the AI domain. For a long time, symbolic AI did not include emotions whatsoever in artificial systems (perhaps following a philosophical idea of emotions as a disturbing or not–possible–to–control element). More recently, embodied (in the sense of an agent or system owning a body; not to be confused with embodiment as bodily responses in other emotion domains) AI has paid attention to the key role of emotions. However, several questions remain unsolved. Several proposals and approaches have been suggested, leading to less or more realistic developments.

There is still a long way to go before developing "fully" emotional systems. Some aspects such as human–machine interaction and adaptive behavior and survival have been partially addressed. For some others, it is still needed to bridge the gap between theory and applications. Issues such as how to use emotional terms in machines or how to develop emotions consisting of components in a human or animal like style (as proposed by Picard) are still under study.

Another aspect to address is the approach followed to develop, let's put it this way, "emotional" system. The framework provided has classified existing research. However, some approaches seem more suitable to come up with a successful physical development (i.e. an architecture, a machine or an autonomous robot) than others. Before any development is considered, it might be worthwhile paying attention to the best approach to follow. Perhaps, the idea will be not so much of developing a system to feel emotions in a human–like way (useless from an engineering viewpoint, I think) but to use them to improve its behavior and adaptation to an environment for a particular goal. Basically, it will be to address first what do we need emotions for in our system in a particular context (and possibly, with a particular goal in mind).

A further concern will be that of terminology used. Some terms related to emotions within the AI domain have been introduced. It seems to exist a partial consensus on the different modeling approaches (as presented in the framework). However, the agreement disappears (or it is less obvious) when classifying the terms and concepts used by researchers (as presented in the AI taxonomy). Why it is so? My personal impression is that most researchers have knowledge of some well–known approaches and developments, choosing one as a starting point (be it either the appraisal theory, the Ortony (OCC) model or Picard's ideas) as more suitable for their purposes. Therefore, the terminology used in the original theory is fully adopted, with some possible further inclusions. Other times, the research is made in an ad–hoc fashion. Likewise, the research seems to be done mostly in isolated research groups without sharing the outcome. As a consequence, a myriad of possible concepts and terms are coined within the AI emotion research without much sharing. It might be worthwhile to try to answer all the questions and issues proposed by [Cañamero, 2005], if we want to come up with a definite solution to the problem of modeling emotions.

When it comes to the interaction between emotion and cognition, once again the starting point followed greatly influences the outcome. Researchers following a so called neuroscience approach focus more on the neural circuitry or networks to be used for emotions. Those following psychology or cognitive science ideas focus more on how emotions affect agents and robot behavior or cognitive processes such as attention or decision–making.

Additionally, it might be necessary to address how to evaluate the benefits gained by incorporating emotion in artificial systems. Which techniques could be used to evaluate the achievement of a better performance is out of the scope of this report. However, it should be pointed out a need for such assessment techniques.

It is my view that only by addressing both the approach to develop the system and the use of an agreed terminology (and underlying concepts) by long–term multi–disciplinary research projects, successful outcomes will be achieved. There is still much work to do on the AI emotion domain.

A last comment on the AI emotion research. The topic is still under progress. Therefore, research continues to address emotions in artificial systems. Only the most relevant (or well–established) and recent works have been presented. Further research is under development, without being possible to evaluate their real application and possible benefits.

Chapter 5

Conclusions and further work

5.1 Conclusions

Some conclusions have been outlined during the different sections. They describe important issues to consider or address within the different domains (psychology, neuroscience and artificial intelligence).

Nevertheless, some general conclusions can be drawn. Although the first aim of the research was to address the approach to emotion and cognition mainly on the AI domain, it was felt that other domains should also be addressed. Several underlying ideas and concepts currently being used in the AI emotion research are heavily based upon theories and concepts belonging to the cognitive psychology and neuroscience. Concepts such as appraisal theory (used in psychology) or circuit models (used in neuroscience) have found their translation when it comes to agents or autonomous robots. See [Cañamero, 2005] for an excellent review.

However, when approaching domains such as psychology and neuroscience, the focus was not so much on the assessment of their benefits or drawbacks. Not being an expert on such domains, the focus was more on the description of main theories and concepts, with the aim of understanding them for their use in the AI domain.

When it comes to the AI domain, the emotion–cognition interaction takes a wicked turn. One should forget about bodily responses, hormones and organs, to consider mere software programs and robotic platforms.

Emotions have recently considered to be useful for the development, behavior and decision–making in agents and autonomous robots. However, the emotion–cognition problem is not addressed in great detail. That such interaction exists has been claimed by several researchers. To tackle how emotion and cognition are intertwined, it is a different story.

Let aside what autonomy means and how it has been implemented. The research tried to address such concept. Surprisingly, not much has been found. Autonomy is understood as two different ideas. On the one hand, to define the capability of robotics platform to move within an environment (the so–

called autonomous robots). On the other hand, to drop hints on the use of emotions for behavior or action selection with the purpose of survival or autonomous response. However, no further accounts on the role of autonomy have been found in the literature. It remains an open issue.

The last conclusion is that of a great remaining work to be done in all the domains. Cognitive psychology has still to understand how emotions are used, what happens when emotions loss control or when emotions are lacking of. The cognitive neuroscience domain has greatly improved with our expanding knowledge of the brain and how it works. New assessment and imaging techniques will help in the pursue.

When it comes to the AI domain, the first work seems to come up with an agreement on what emotion means and their benefits. As stated in the AI section, researchers within the domain seem far from consensus. A myriad of terms and concepts used in the domain do not ease the task of modeling artificial emotions. An additional issue to address is the anthropomorphic approach currently used. A deeper analysis of emotions really useful for robots and agents is needed. Further conceptualization should be done to analyze the influence of emotions not only at low levels of cognition, but at higher and not always so well known levels.

A final statement: there is no doubt that the AI domain has overcome the tradition of regarding emotions as sub–products of our embodiment to consider emotions as widespread elements of cognition, behavior and intelligence.

5.2 Further work

This research is part of a multidisciplinary and long–term european project, ICEA (Integrating Cognition, Emotion and Autonomy) project that will develop the first cognitive systems architecture integrating cognition, emotion and autonomy (bioregulation and self-maintenance), based on the architecture and physiology of the mammalian brain.

Bearing it in mind, it would have been interesting to consider the past and current research of the groups participating in the project. The purpose will be twofold. Firstly, to include their terminology and concepts both in the framework and the taxonomy to ensure a common use of lexicon and core ideas. Secondly, to assess how the interaction emotion and cognition has been addressed by such research groups. However, the time and distance constraints have not allowed such analysis. It remains to be done.

A second comment on research to accomplish. The outcome of the ICEA project are two central, integrated platforms, rat-like in appearance, perceptual, and behavioural capacities. The ICEAbot robot platform, equipped with multimodal sensory systems will serve as a real-world testbed and demonstrator of the behavioural and cognitive capacities derived from models of rat biology. A 3-D robot simulator, ICEAsim, based on the physical ICEAbot, will demonstrate the potential of the ICEA architecture to go beyond the rat model and support cognitive capacities such as abstraction, feelings, imagination, and planning. ICEAsim will serve as the main platform for exchange and technical integration of models developed in different parts of the project.

Therefore, a wider research on past or ongoing research projects within the AI community should be considered. The particular research presented in the current report has attempted to review some of the most significant projects within the Artificial Intelligence domain. However, the rest of research groups belonging to ICEA could provide useful insights on their knowledge on different projects. Even so, it would be an herculean task to consider every single project already developed or currently under development. Nevertheless, the more we know, the better.

A last comment regarding the taxonomy. The idea was to show the terms currently being used by the AI research community. However, it is far from being complete. Once the ICEA research members agree upon a set of common terms, a more detailed taxonomy could be developed.

Bibliography

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

[Cannon, 1927] Cannon, W. (1927). The james–lange theory of emotion: a critical examination and an alternative theory. *American Journal of Psychology*, 39:106–124.

[Carter and Pasqualini, 2004] Carter, S. and Pasqualini, M. S. (2004). Stronger autonomic response accompanies better learning: a test of damasio's somatic marker hypothesis. *Cognition and Emotion*, 18(7):901–911.

[Cañamero, 1997] Cañamero, L. (1997). Modeling motivations and emotions as a basis for intelligent behavior. In Johnson, W. L., editor, *First International Conference on Autonomous Agents*, pages 148–155. ACM Press.

[Cañamero, 2000] Cañamero, L. (2000). Designing emotions for activity selection. Technical Report TR–DAIMI PB 546, University of Arhaus.

[Cañamero, 2001] Cañamero, L. (2001). Emotions and adaptation in autonomous agents: a design perspective. *Cybernetics and Systems: An International Journal*, 32:507–529.

[Cañamero, 2005] Cañamero, L. (2005). Emotion understanding from the perspective of autonomous robots research. *Neural Networks*, 18(4):445–455.

[Cañamero and Fredslund, 2000] Cañamero, L. and Fredslund, J. (2000). How does it feel? emotional interaction with a humanoid lego robot. In Dautenhahn, K., editor, *Socially Intelligent Agents: The Human in the Loop*, AAAI 2000 Fall Symposium. AAAI Press.

[Cañamero and Gaussier, 2005] Cañamero, L. and Gaussier, P. (2005). Emotion understanding: robots as tools and models. In Nadel, J. and Muir, D., editors, *Emotional Development: Recent Research Advances*. Oxford University Press.

[Charland, 2005] Charland, L. (2005). Emotion experience and the indeterminacy of valence. In Barrett, L. F., Niedenthal, P. M., and Winkielman, P., editors, *Emotion and Consciousness*, pages 231–254. The Guilford Press.

[Custódio et al., 1999] Custódio, L., Ventura, R., and Pinto-Ferreira, C. (1999). Artificial emotions and emotion–based control systems. In *7th IEEE International Conference on Emerging Technologies and Factory Automation (EFTA'99)*, Barcelona, Spain.

[Damasio, 1994] Damasio, A. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*. Putman.

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

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

[Fellous and LeDoux, 2005] Fellous, J.-M. and LeDoux, J. (2005). Toward basic principles for emotional processing: what the fearful brain tells the robot. In Fellous, J.-M. and LeDoux, J., editors, *Who needs emotions: the brain meets the robot*, pages 79–116. Oxford University Press.

[Friedenberg and Silverman, 2006] Friedenberg, J. and Silverman, G. (2006). *Cognitive science: an introduction to the study of mind*. Sage Publications.

[Frijda, 1986] Frijda, N. (1986). *The Emotions*. Cambridge University Press.

[Gadanho and Custódio, 2002] Gadanho, S. and Custódio, L. (2002). Aysnchronous learning by emotions and cognition. In *7th International Conference in Simulation of Adaptive Behavior (SAB'02)*, Edimburgh, UK.

[Gadanho and Hallam, 1998] Gadanho, S. and Hallam, J. (1998). Exploring the role of emotions in autonomous robot learning. In *Emotional and intelligent: the tangled knot of cognition*, pages 84–89. D. Cañamero.

[Gadanho and Hallam, 2001] Gadanho, S. and Hallam, J. (2001). Emotion–triggered learning in autonomous robot control. *Cybernetics and Systems: an international journal*, 32:531–559.

[Gray et al., 2002] Gray, J., Braver, T., and Raichie, M. (2002). Integration of emotion and cognition in the lateral prefrontal cortex. In *PNAS 2002*, volume 99, pages 4115–4120.

[Gray et al., 2005] Gray, J., Schaefer, A., Braver, T., and Most, S. (2005). Affect and the resolution of cognitive control dilemmas. In Barrett, L. F., Niedenthal, P. M., and Winkielman, P., editors, *Emotion and Consciousness*, pages 67–94. The Guilford Press.

[Hudlicka, 2004] Hudlicka, E. (2004). Beyond cognition: modeling emotion in cognitive architectures. In *sixth International Conference in Cognitive Modeling, ICCM'04*, pages 118–123.

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

[Lange, 1885] Lange, C. (1885). The emotions. In Dunlap, E., editor, *The Emotions*. Williams and Wilkins, Baltimore, 1922 edn edition.

[Lazarus, 1991] Lazarus, R. (1991). *Emotion and Adaptation*. Oxford University Press, New York.

[LeDoux, 2000] LeDoux, J. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, 23:155–184.

[LeDoux, 2002] LeDoux, J. (2002). *Synaptic self: how our brains become who we are*. Penguin.

[Niedenthal et al., 2005] Niedenthal, P., Barsalou, L., Ric, F., and Krauth-Gruber, S. (2005). Embodiment in the acquisition and use of emotion knowledge. In Barrett, L. F., Niedenthal, P. M., and Winkielman, P., editors, *Emotion and Consciousness*, pages 21–50. The Guilford Press.

[Oatley and Jenkins, 1996] Oatley, K. and Jenkins, J. (1996). *Understanding Emotions*. Blackwell, Cambridge, MA.

[Oatley and Johnson-Laird, 1987] Oatley, K. and Johnson-Laird, P. (1987). Towards a cognitive theory of emotions. *Cognition and Emotion*, 1:29–50.

[of Duke,] of Duke, U. Emotion research.

[Oliveira and Sarmento, 2003] Oliveira, E. and Sarmento, L. M. (2003). Emotional advantage for adaptability and autonomy. In *The Second International Joint Conference on Autonomous Agents & Multiagent Systems, AAMAS 2003*, pages 305–312. ACM.

[Ortony et al., 1988] Ortony, A., Clore, G., and Collins, A. (1988). *The cognitive structure of emotions*. Cambridge University Press.

[Pfeifer, 1987] Pfeifer, R. (1987). Artificial intelligence models of emotion. In Hamilton, V., Bower, G., and Frijda, N., editors, *Cognitive Perspectives on emotion and motivation*, NATO ASI, pages 287–320. Kluwer Academic Publishers.

[Phelps, 2005] Phelps, E. (2005). The interaction of emotion and cognition: insights from studies of the human amygdala. In Barrett, L. F., Niedenthal, P. M., and Winkielman, P., editors, *Emotion and Consciousness*, pages 51–66. The Guilford Press.

[Phelps and LeDoux, 2005] Phelps, E. and LeDoux, J. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron*, 48(2):175–187.

[Prinz, 2005] Prinz, J. (2005). Emotions, embodiment and awareness. In Barrett, L. F., Niedenthal, P. M., and Winkielman, P., editors, *Emotion and Consciousness*, pages 363–384. The Guilford Press.

[Rolls, 2005a] Rolls, E. (2005a). What are emotions, why do we have emotions and what is their computational basis in the brain. In Fellous, J.-M. and LeDoux, J., editors, *Who needs emotions: the brain meets the robot*, pages 117–146. Oxford University Press.

[Rolls, 2005b] Rolls, E. T. (2005b). *Emotion explained*. Series in Affective Science. Oxford University Press.

[Sarmento, 2004] Sarmento, L. M. (2004). *An Emotion–based Agent Architecture*. PhD thesis, Departamento de Ciência de Computadores, Faculdade de Ciências da Universidade do Porto.

[Schachter and Singer, 1962] Schachter, S. and Singer, J. (1962). Cognitive, social and physiological determinants of emotional state. *Psychological Review*, 69:379–399.

[Scherer, 1999] Scherer, K. (1999). Appraisal theory. In Dalgleish, T. and Power, M., editors, *Handbook of Cognition and Emotion*, chapter 30, pages 637–663. Wiley, New York.

[Scherer, 2005] Scherer, K. (2005). Unconscious processes in emotions: the bulk of the iceberg. In Barrett, L. F., Niedenthal, P. M., and Winkielman, P., editors, *Emotion and Consciousness*, pages 312–334. The Guilford Press.

[Sloman, 1982] Sloman, A. (1982). Towards a grammar of emotions. *New Universities Quaterly*, 36(3):230–238.

[Sloman, 2001] Sloman, A. (2001). Beyond shallow models of emotion. *Cognitive Processing*, 2(1):177–198.

[Sloman, 2004] Sloman, A. (2004). What are emotions theories about. In *Architectures for Modelling Emotions*, AAAI Spring Symposium, Stanford.

[Sloman, 2005] Sloman, A. (2005). Do machines, natural or artificial, really need emotions? Utrecht. NWO Cognition Programme.

[Velásquez, 1997] Velásquez, J. (1997). Cathexis: a computational model of emotions. In *International Conference on Autonomous Agents*.

[Velásquez, 1998] Velásquez, J. (1998). A computational framework for emotion–based control. In *Grounding emotions in adaptive systems, SAB'98*, Zurich, Switzerland.

[Ventura and Pinto-Ferreira, 1999] Ventura, R. and Pinto-Ferreira, C. (1999). Emotion–based agents: three approaches to implmentation. In Velasquez, J. D., editor, *Workshop on Emotion–based agent architectures*, Seattle, U.S.A.

[Winkielman et al., 2005] Winkielman, P., Berridge, K., and Wilbarger, J. (2005). Emotion, behavior and conscious experience: once more without feeling. In Barrett, L. F., Niedenthal, P. M., and Winkielman, P., editors, *Emotion and Consciousness*, pages 335–362. The Guilford Press.

Document information

Reference: ASLab-ICEA-R-2006-001 v 1.0 Draft
Title: State of the Art on Emotion
Subtitle: Integrating Cognition and Emotion for Autonomous Systems
URL: http://www.aslab.org/documents/ASLab-ICEA-2006-001.pdf
Date: 2006-04-15

EUROPEAN INTEGRATED PROJECT IST-027819
Integrating Cognition, Emotion and Autonomy

www.ist-icea.org