FP6-511931

Mind RACES

from Reactive to Anticipatory Cognitive Embodied Systems

DELIVERABLE D6 (D5.1)

Critical Comparison of Architectures Modelling Motivation and Emotions

Due date of deliverable:
31 / 05 / 2005

Actual submission date:
13 / 07 / 2005

Start date of project:
01 / 10 / 2004

Duration:
36 month

Organization name of lead contractor for this deliverable:
IST

Revision:
3

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	**X**
PP	Restricted to other programmes participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Document identifier:	**DELIVERABLE_WP5_N_1**
Date:	**13/07/2005**
Work package:	**WP5**
Partner(s):	*IDSIA, IST, ISTC-CNR, LUCS, NBU, OFAI, UW-COGSCI, NOZE*
Lead Partner:	*IST*
Document status:	*Approved*
Deliverable identifier:	**WP5_D5.1**

Delivery Slip

	Name	**Partner**	**Date**	**Signature**
From	JOÃO DIAS, CARLOS MARTINHO, ANA PAIVA, & CRISTIANO CASTELFRANCHI, MARIA MICELI	IST, ISTC-CNR	28/06/2005	
Verified	RINO FALCONE AND CRISTIANO CASTELFRANCHI	ISTC-CNR	05/07/2005	
Approved by	RINO FALCONE	ISTC-CNR	08/07/2005	

Files

Software Products	**User files**
MS Word™	D5_1.DOC

Project information

Project acronym:	Mind Races
Project full title:	MIND RACES: from Reactive to Anticipatory Cognitive Embodied Systems
Proposal/Contract no.:	IST-511931
Project Manager:	ISTC_CNR
Name:	Rino Falcone
Address:	CNR-ISTC via S. Martino della Battaglia,44 00185 Rome ITALY
Phone:	+39 06 44 595 253
Fax:	Fax: +39 06 44 595 243
Mobile:	
E-mail	rino.falcone@istc.cnr.it

TABLE OF CONTENTS

Summary

This deliverable provides an outline of the reciprocal relations between emotion and anticipation, with a strong emphasis on emotion.

We begin by presenting the current state of the study of emotion in three fields: philosophy of mind, affective neuroscience, and psychology.

We then select a few representative architectures and application models from the field of affective computing, and presented them in some detail. From our research, it is clear that most of the current systems do not deal with the concept of anticipation explicitly, although most support planning capabilities. Anticipation is a novel approach to Affective Computing that (at the least) will provide with a valuable fresh insight on affective architectures, that appear to have somewhat stagnated in a pragmatic shallowness.

Afterwards, we present the concept of an anticipatory system, and briefly explained its value when confronted with the universality of reactive systems. Finally, two affective anticipatory approaches are presented. The former approach is a high-level more formal approach, oriented towards the integration of anticipatory-based emotions in BDI models. The latter is a sub-symbolic lower level approach. Both of them are two novel anticipation-based approaches in the field of Affective Computing.

Introduction

Emotion

Our everyday decisions are mediated by emotions. If I am walking in the woods, and I see a bear appear on the path ahead of me, my muscles tense, my heart races. I feel afraid. I tense in readiness to run away. Why? I could argue: I have experiences, and as a result, my autonomic nervous system creates physiological events such as muscular tension and heart rate increase. Although there is still some on-going uncertainty where exactly "emotions" enter the physiological loop, one thing is certain: emotions are present in every decision we make; bodily sensations condition us for action, emotions grab our attention and influence at least some lower cognitive processing.

Anticipation

Our everyday behaviour is based on the implicit employment of predictive models. If I am walking in the woods, and I see a bear appear on the path ahead of me, I will immediately tend to run away. Why? I could argue: because I can *foresee* a variety of unpleasant consequences arising from failing to do so. The stimulus of the action is not *just* the sight of the bear, but rather the output of the model through which I predict the consequences of direct interaction with the bear. I thus change my present course of action, in accordance with my model prediction. Or, put another way, my behaviour is not simply *reactive* but rather is *anticipatory*.

Emotion and Anticipation

As the revisited example of William James (James, 1884) in the light of new evidence from the study of emotion can forecast, the relationship between emotion and anticipation is narrow and manifold. To start with, one of the functions ascribed to emotions is precisely that of anticipating events, especially when they are relevant to central concerns and the well-being of the organism. In fact, the functional value of emotions has been widely acknowledged in the last decades, and their anticipatory, interpretive and evaluative features have been especially emphasized (e.g. Fridja, 1986; Oatley and Jenkins, 1996; Parrot and Schulkin, 1993; Smith and Lazarus, 1990). Thus the negative bias against emotions, and more generally "irrational" responses, traditionally viewed as contrary to utility and disruptive for both rational thinking and effective behaviour, has been practically reversed.

This document will try to provide with a systematic outline of the reciprocal relations between emotion and anticipation, with a strong focus on emotion. We will first address the concept of emotion and present some relevant architecture of current affective systems, showing that most affective system do not deal with anticipation explicitly. Then, we will address the concept of anticipation and show how emotions are an intrinsic characteristic of such an approach. Afterwards, we will resume the debate among partners regarding how anticipatory affective systems differ from "normal" anticipatory system, and how such differences can be tested. Finally, we will present two cognitive approaches of anticipatory affective systems, a formal higher-level approach, and a sub-symbolic lower level approach: two novel anticipation-based approaches in the field of Affective Computing.

2. Emotion and Cognition

Every individual has experienced emotion at some point. Emotion means movement, something that moves us: the body that is physically aroused during emotion (e.g. pounding heart, sweating palm); our motivation to take action, moved by emotions such as fear, anger. Emotion is a complex process that differs from other cognitive processes. It can encompass thoughts about a stimulus, plans about the future concerning the stimulus, and physiological changes associated with the stimulus with which it is associated. Simply being reminded of a stimulus, either consciously or unconsciously, can elicit the subjective experience of emotion.
We will briefly review how Emotions are perceived in three different fields: Philosophy of Mind, Affective Neuroscience and Psychology.

Philosophy of mind

In contemporary philosophy of mind, mental phenomena are usually divided in two broad categories: the *intentional* and the *phenomenal*. The intentional aspect of the mind is especially evident in propositional attitudes (e.g. beliefs, desires…), the states that carry representational content. The phenomenal aspect consists in the subjectively felt qualitative aspect of consciousness, whose dimensions are called *qualia*.

Emotions do not easily fit in one or the other category; rather they are a mixed case. Emotions have a representation context (I am afraid *of* the bear), while undeniably have a phenomenal aspect (what it *is* to be afraid). The phenomenal aspect, in particular, is believed by many to constitute the real challenge of all (Chalmers, 1996). If both aspects of the mind could be figured out, then emotions would be much easier to understand. Thus, most of the philosophers of mind reach the paradoxical conclusion that to better understand emotions, they are better off not studying them.

More recently, however, emotions have been the object of a renewal of attention from psychologists and neuroscientist, mainly to some recent developments in the neuroanatomy of emotions.

Affective neuroscience

Affective neuroscience is the name given by Jaak Panksepp to the discipline whose goal is to provide a “neurological understanding of the basic emotional operation systems of the mammalian brain and the various conscious and unconscious internal states they generate” (Panksepp, 1998).
Researchers began expressing interest in the anatomy of emotion at the turn of the twentieth century. Experimenting with cats, a series of studies beginning with Philip Bard’s led to the beliefs that the thalamus and the hypothalamus were the neural mechanisms likely to be responsible for both overt expression of emotion and the autonomic responses associated with emotional states. The cortex was believed to inhibit the thalamus and hypothalamus, in the sense that reason can supplant the “will” of emotions.

More recently, pre-eminent authors as Jaak Panksepp (Panksep, 1998), Joseph Ledoux (Ledoux, 1996) and António Damasio (Damasio, 1994) provided with new evidence about the neurophysiology of emotions. Although we are still far from a comprehensive understanding of emotions, we certainly know a lot more about the neurophysiology of emotion than a century ago, although the phenomenal aspect did not evolve much.

In the light of contemporary neuroscience, Emotion is characterized in three general themes:

The concept of emotions is mainly biological, especially, neurobiological.
"Emotion is the combination of a mental evaluative process, simple or complex, with dispositional responses to that process, mostly toward the body proper, resulting in an emotional body state, but also toward the brain itself, resulting in additional mental changes." (Damasio, 1994)

Emotions are a product of evolution, and exist because of the survival function they fulfil.
"Emotions did not evolve of conscious feelings. They evolved as behavioural and physiological specializations, bodily responses controlled by the brain that allowed ancestral organisms to survive in hostile environments and procreate." (Ledoux, 1996)

Emotions do have a phenomenal aspect, a function of natural selection.
Neuroscientists mainly differ among themselves in regard to the importance of the phenomenal aspect in the definition of emotion. As an example here are the views of the three mentioned authors. Panksepp considers it an essential component. Ledoux states it as an optional product of a non-emotional mechanisms present in Humans. Damasio describes it as a characteristic of high-order emotions but not primary emotions. The essence of emotion is the collection of changes in body state. Damasio insists not so much on the nervous system than on the entire body.
Contemporary affective neuroscience has explicitly demystified several sins in the study of emotions (Davidson, 2003), sustaining that:
Emotions involve important peripheral and visceral components that are crucial in understanding their nature. Emotions are not only "in the head".
There is an overlap in the circuitry involved in cognitive and affective processing, and not separate and independent neural circuits. Further, affect is both sub-cortical and cortical.
A claim often supporting the notion of "basic emotion" is that specific emotions are instantiated in discrete locations in the brain. When a stimulus is used to arouse emotion in humans or in animals that have a fairly complex brain, it is important not to assume that a single process has been activated. Emotions is comprised of many sub-components and is best understood not as a single monolithic process but rather as a set of differentiated subcomponents that are instantiated in a distributed network of cortical and sub-cortical circuits. The subcomponents that get triggered vary as a function of many different processes, including the nature of the elicitor, and the context in which the emotion gets elicited. The neuro-imaging literature is replete with demonstrations that affective stimuli activate broad network of both sub-cortical and cortical regions.

Emotions are different in structure across age and species.
If the neural substrate of emotions gleaned from the study of rodents (Panksepp, 1998) has been crucial for progress in the field, the differences with human anatomy imply

differences in the nature, function, and complexity of emotions across species. Another unclear assumption is that the same basic incentive conditions will elicit the same basic emotional process in individuals at different ages.

Emotions are not only conscious feeling states.
Most of the psychological literature of emotion assumes that emotions are conscious feeling states. A vast number of studies depend upon self-report measures to make inferences about the presence of emotional states. Failure to find detectable changes on self-report measures is sometimes offered as evidence that emotion was not activated, and conversely the presence of self-reported emotion is taken as evidence that emotion has been activated. Although the experiential side of emotion is unquestionably important, it is also clear that much of the affect we generate is likely to be non-conscious (Damasio, 1994).

Emotions cannot be studied from one perspective only.
Although researchers are now able to identify the regions of the brain and neurological systems and anatomical bases that support the theory of cognition and emotion in decision making, the challenge that faces the study of emotion in Affective Neuroscience is similar to that once faced by investigators studying cognition - the decomposition of complex emotional phenomena into more elementary components.
The three following theories illustrate the current beliefs and directions in the studies of emotion in neuroscience. Through them, a basic idea is present in all modern physiological views of emotions: that emotion and cognition are inextricably tied and thought to overlap neural systems.

Somatic marker hypothesis

The most cited theory used to support emotion in decision making is the somatic marker hypothesis. The somatic marker hypothesis theory of decision-making is a systems-level, neuroanatomical, and cognitive framework influenced by emotion (Bechara, Damasio and Damasio, 2000). This theory was the result of investigating the link between the abnormalities in emotion and feeling of the patient with brain damage and their severe impairment in judgement and decision-making in real-life.
Damasio postulated the somatic marker hypothesis based on the theory that emotion occurs as a response to a stimulus; we interpret the physiological changes we experience when presented with a stimulus as a particular emotion. According to Damasio (Damasio, 1994), emotions involve both the neural structures that represent body states and the structures that link the perception of external stimuli to body states. Somatic markers, or bodily changes, are linked to external events and work together to influence cognitive processing, both conscious and unconscious. Emotion is fundamental to the survival of an individual in the environment, particularly in social settings for humans. Finally, emotion is not only a fundamental experience for all higher animals but also a necessary experience for individuals to make rational decisions: individuals with emotional impairments exhibit an impaired ability to make personal or social decisions. One particular relevant example is the experiment Bechara (Bechara, 2004) conducted using the Iowa gambling test: the results revealed that the normal patients eventually avoided decks with no reward but favoured decks that gave a positive outcome; in contrast, the damaged patients continued to use faulty decks even though they received no rewards, as if unable to anticipate the outcome of their decisions.

Cognitive emotional interactions

LeDoux postulated a theory of cognitive-emotional interactions. According to this theory (focused in fear) the detection of fear is inherent and the neural mechanism underlying it produces a particular subjective feeling that is interpreted as fear. As humans, we experience both unconscious fear responses as well as conscious feelings of fear. LeDoux studied fear through conditioning and demonstrated that damage in the amygdala resulted in an impairment in fear conditioning. LeDoux assumes that there are two separate neural networks at work in fear conditioning: one resulting from evolution (e.g. seeing a snake for the first time) and the other from learning (e.g. seeing a doll, after having watched the negative effects of a bomb-trapped toy in a prior situation). A stimulus is interpreted as dangerous and then the attention is directed to it. This is important because it also serves to explain why a stimulus can be perceived as threatening in one situation but not in another. These structures work together to determine if a fear response is necessary based on both evolutionary conditioning for fear as well as learned fear conditioning.

The neural bases involved in this process consist of both a cortical and a sub-cortical route. When emotional processing is occurring, the thalamus activates the amygdala (the sub-cortical route) and the cortex at slightly different times. The stimulus activates the amygdala first, which allows emotional responses to be produced before an individual is conscious of it. For example, if I hear a rustling in bushes while walking through the woods, there might not be time to evaluate the situation, as the cause of noise might be a real threat. The amygdala acts on this information and produces an anticipatory reflex-like response (e.g. running), which may help to protect the individual from impending danger. Shortly after the amygdala responds, the stimulus travels down the second route for emotional learning and reaches its destination: the cortex. This structure evaluates the stimulus and either inhibits or facilitates an emotional response. That is, it will allow the individual to determine that the cause of the noise is a bear and will facilitate running. Conversely, if it determines the cause to be a squirrel, the emotional response will cease (LeDoux, 1996).

Cognitive Asymmetry and Emotion

According to the cognitive asymmetry and emotion perspective, since emotion is related to cognition and cognition is highly asymmetrical, emotional systems should also be asymmetrical. Hence, laterality studies, which examine differences between functions of the hemispheres, have been conducted in the search for anatomical bases for emotion (Bechara, 2004).

Gainotti found that patients who had lesions to the left hemisphere exhibited catastrophic reactions such as fearfulness and depression, while those with right hemisphere lesions displayed indifference. It can be assumed that these lesions reveal the true nature of the hemisphere that is still intact, thus adding more support for the contention that the right hemisphere holds a negative and pessimistic emotional view of the world, responsible for the production of strong emotions such as fear and anger, as opposed to the more positive view of the left hemisphere. Further, Gainotti postulates that the right hemisphere is assumed to control the automatic component of emotion, while the left hemisphere controls the overall cognitive control of emotions.

Other lateralization studies have examined differences in emotional processing in patients who have sustained brain injuries. Studies on mood have revealed that left hemisphere lesions, mainly when located in the frontal lobe, cause a flattening of mood and sometimes depression, especially when language is affected. Facial expression has also been examined. Generally, it has been found that when patients sustain lesions to the more anterior areas of the hemisphere, there is a reduction in frequency and intensity of facial expression, and studies by Campbell, Moscovitch and Olds have demonstrated that facial expressions conveying emotional states predominantly occur on the left side of the face. Also, evidence has been found that the right hemisphere is specialized for perceived facial expressions. Laterality studies examining audition have produced analogous results to those of vision studies: the left ear was specialized in identifying emotional tone, while the right ear had an advantage in identifying content. Prosody and content of language also appear to be specialized according to hemisphere. The left hemisphere is responsible for content and the right hemisphere appears to play a role in the tone of voice used in language. Patients with right hemisphere lesions displayed aprosody, in which they read sentences with flat affect as compared to those with left hemisphere lesions. Two examples of these aprosodias are motor aprosodia (an inability to produce affective components of language and results from damage to Broca's area) and sensory aprosodia (a deficit in the interpretation of emotional components of language and results from damage to Wernicke's area). All these studies are particularly relevant when considering models of the expressive component of emotion, as facial expression and emotional speech, both in terms of generation and recognition.

Psychology of Emotion

We are far from being able to provide an answer to William James question "what is an emotion" (James, 1884). Young stated that "almost everyone except the psychologist knows what an emotion is" (Young, 1973). Indeed, each psychologist focuses on different features of emotion and defines emotion differently, according to their approach. As a result, Psychology provides us with a plethora of definitions for emotion. In Psychology, Lazarus (Lazarus, 1991), Oatley and Johnson-Laird (Oatley and Johnson-Laird, 1987) provided with impressionably complete theories and built a bridge between academic and popular psychology, but other theories must not go unnoticed as they approached other themes: Mandler (Mandler, 1990) approached the subject of consciousness in emotion; Panksepp's connection with physiology (Panksepp, 1998); Plutchik context of emotion in terms of evolution (Plutchik, 1991); Eckman approach of emotion and facial expression (Ekman, 1982); Izard in term of the distinction between emotion and cognition (Izard, 1992); Malatesta-Magai (Malatesta-Magai, Izard and Camras, 1991) and Lewis (Lewis, 1993) in terms of the development aspect; are but to site a few.

History of the Psychology of Emotion

The theories of emotion have their root in philosophy. In the end of the 19th century, the psychologist and other researchers began expressing an interest for emotion. The *first theories* of Emotion (Darwin, 1872; James-Lange, 1884; Cannon-Bard, 1927) focused on the origins and the development of emotion. They distinguished between what is an emotion and what is not. Right from the start, they considered the phenomenal aspect of emotion (what it is to feel an emotion) as having an important role. Although recognizing that emotions would also have a behavioural and expressive aspect (a critical function within the evolution perspective), these theories researched where would the emotions be located, and started the investigation on the physiology of emotion. These first theories also started the first debates regarding whether emotion should be based on biology or social interaction, and showed how difficult it would be to handle emotion without a cognitive component.

1. Phenomenologist Theories

The *phenomenology* of emotion deals essentially with the nature of the emotional experience. They are interested in what people feel right here right now (as a whole), and how these experiences can be manipulated by the control of previous experiences. By emphasizing the subjective aspects of emotions, this theoretical perspective has much in common with the popular approach regarding emotions:. But although the subjective aspect of emotion is not always considered scientifically relevant, it cannot be denied or hidden.

As the phenomenologist theories do not speak much of the physiology or the cognitive aspects of emotions, they are inherently incomplete. However, they have their own merit. Based on the popular beliefs regarding emotions, they are heuristically acceptable and provide with definitions for discrete emotions that define their context in the overall function of the body, and that can be applied in therapeutics.
One particularly interesting example is the work of Rivera and his "emotional climate" (Rivera, 1977) that argues that emotions are not particular of a person, but are created between them.

2. Behavioural Theories

As opposed to the phenomenologist approach, the *behavioural* approach to emotion departs from the subjective experience of emotion and focuses on what can be directly observed or measured. Behaviourists do not deny the existence of emotional sensations, but they shift the focus of the research on what is supposed to be more adequate for conventional science methodologies. They support the vision that emotions are related with changes in behaviour, but consider the idea of an *emotional state.* They generally leave out all cognitive aspects of emotion[1]. As such, this approach is rather restrictive.

The strength of behaviourist theories is that they are linear and provide with predictions that can be easily tested. Behaviourists are good at characterizing discrete emotions, giving them a motivational structure, at handling the creation and developmental aspects of emotions, describing their general effect in the overall performance of man and pointing some therapeutic application.

The weakness of behaviourist theories is that they are generally very narrow and do not have a great heuristic value (excepting some theories as Gray, 1987). To distinguish between emotion and non-emotion under the behaviourist view is a delicate matter. Complex emotions such as guilt, shame or envy, are left out of the behaviourist range. They do not say very much about the unconscious part of emotion. Most do not provide with biological or socio-cultural connections, primarily due to its conceptual distancing to the popular approach of emotion. This is surprising as emotion is a phenomenon

[1] Although, ironically, most researchers working exclusively in behaviourism end up concluding that it is impossible to work solely in terms of behaviourism without considering cognition.

predominantly social, and that the social aspects of emotion could also be investigated using the behavioural approach.

3. Physiological Theories

Physiological theories support that emotions have a neurophysiologic basis and are of biological importance in the history of evolution. The physiology of emotion has been present since Psychologists started studying emotions, as it is clear that the neurophysiologic reactions of the body are involved in some way in the process of emotion: we feel them in ourselves and perceive them in others.
The border between what is an emotion and what is not is rather thin in physiological approaches, as the concept of activation is not clearly defined. Is the physiological activation global or restricted to certain areas? Does this activation constitute the emotion? Is it independent or a dependent measure of another process? Is the activation the process that leads to the emotion or the reverse? Maybe emotions are all that.

Physiological theories are more focused on the generation of emotion than on its development: although they refer the impact of the emotion on overall function of the body, Psychologists will tend to agree that they still have a low impact in therapy. One important fact, however, is that all physiological theories of emotion are related with cognition: they explore the connection between emotion and other phenomena as cognition and motivation, although not considering very deeply appraisal, consciousness, social or inter-personal matters. Plutchik (Plutchik, 1991), Panksepp (Panksepp, 1998) and Scherer (Scherer) are example of physiological approaches that have good heuristic value.

4. Cognitive Theories

Cognitive theories of emotion deal essentially with the relation between emotion and cognition. These theories approach lightly the behavioural and the physiological components of emotion, generally do not embrace the phenomenology nor the socio-cultural aspects of emotion, barely explain its developmental aspects or biological origins, and rarely distinguish between emotion and non-emotion. So, why are these theories so popular?
Cognitive theories are very stimulating as they try to "explain" emotion and have a very good heuristic value, even if testing such theories is no trivial matter. Furthermore, and somewhat related to the ascension of cognitive psychology, cognition has ended up finding a place in most theories of emotions, independently from the original orientation (behavioural, physiological…) and, as such, acquired a central role in the study of emotion.
The first cognitive theories of emotion (e.g. Schachter) emphasised the concept of emotion as composed of two factors: physiology and cognition. Far from the more complex theories (e.g. Oatley, Johnson-Laird, Lazarus), even the first cognitive theories explained the relation of perception and action in the process of emotion, an approach transposable to agent-based computational models (Minsky, 1986). Hence, cognitive models are popular within the field Affective Computing (Picard, 1997). Many subjects have, since then, been approached by the cognitive field of emotion as information processing, network models, consideration of goals and plans.
However, some themes do appear recurrently: all cognitive approaches on emotion share serious concern regarding the process of *appraisal* (e.g. Fridja, 1986; Ortony, Clore and Collins, 1988). Arnold theory (Arnold, 1970) was one of the first to emphasise the importance of appraisal, a tradition that has been continued by Ellsworth

(Ellsworth, 1991). These theories do not only accept the fact that such a process exists, but that it is an integral part of emotion, and relate this appraisal structure with the discrete nature of emotion. Furthermore, most of the more recent theories, besides increasing the relevance of appraisal, approach the concept of *signification* and integrate emotion in a much broader perspective, growing to a point where the theories become systems theories, in the sense of how emotion interacts with other psychological systems.

5. Developmental Theories

The first developmental theories of emotions were elaborated in the attempt to provide an account for the emotional development and difference of emotional processes at different ages. However, only more recently (e.g. Izard, 1992; Malatesta-Magai, 1991) gained a more consistent form. Most of them do consider cognition, motivation and appraisal. They mention the possible biological origins of emotion and its subsequent social development, how it evolves and relates with personality, all from the developmental point of view.

Current developmental theories provide with a convincing summary of existing data, deal with the matter of discrete emotion and the inter-dependent. They provide with an account for how the emotional aptitudes are acquired and change over time, and how, in reverse, emotions condition development. The relation of emotion with moral is also considered. Older development theories considered emotions as initial instinctive motors, interacting with cognition but not inter-connected with it. More recent theories (e.g. Fisher, 1990) conferred to cognition (and mainly to appraisal) a central role, and have increased their distance from the seminal work of Watson.

6. Social Theories

Most theories based on social psychology are vague regarding what they mean by "emotion is primarily a social phenomenon". They are mainly focused on the inter-personal communication of emotions. Most do confine to facial expression and recognition, mainly influenced by the work of Ekman (Ekman, 1982), even if they rarely discuss what is non-emotional versus emotional in facial expression. Besides the work of Ekman, few theories approach the physiological or biological aspects of emotion, but as happened with developmental theories, most do relate to cognition and appraisal.

There is distinctive motor behaviour associated with emotion. This refers to tone of voice, posture, and facial expressions. These behaviours are important in conveying emotional states, as they often speak louder than words. For example, although an individual might verbally indicate a state of happiness, one's perception is altered when accompanied by the motor behaviour of uncontrollable sobbing. Furthermore, research as shown that facial expression indicative of six specific emotions is recognized across culture, thus supporting the universal inclination for humans to physically convey their emotions.

An important aspect of social theories is the establishment of emotional phylogeny. The maintenance of certain emotional features across species is ignored by many researchers, even if they do contextualize emotions under an evolutionary perspective, and give relevance for the study of emotion in primates. Another important aspect is how emotions structure relationships. Strangely, more interesting contributions to the social aspect of emotion came from outside psychology (e.g. cultural studies).

7. Clinical Theories

By definition, emotion is involved in all affective perturbations. Clinical theories deal with specific pathologies: anxiety, depression and generally do not provide with a broad view of emotion. Again, cognition and appraisal play a fundamental part in these theories, and guide therapy strategies. One of the major contributes of such theories is the concept of *uncertainty* in the study of anxiety.

Defining Emotion

Any attempt to provide with a formal definition of emotion should try to include all the significant aspects of emotions, while attempting to differentiate it from other psychological processes. This is what Kleinginna (Kleinginna, 1981) did, after an extensive review of the literature, suggesting the following definition:

Emotion is a complex set of interactions among subjective and objective factors, mediated by neural/hormonal systems, which can:

Give rise to affective experiences such as feeling of arousal, pleasure/displeasure.

Generate cognitive processes such as emotionally relevant perceptual effects, appraisals, labelling processes.

Active widespread physiological adjustments to the arousing conditions.

Lead to behaviour that is often, but not always, expressive, goal-directed and adaptive.

This definition seems to cover all the relevant aspects of emotions. It includes the two dimensions identified by the philosophers of mind: the phenomenal aspect is described in (1), and the intentional aspect in (2) and (4). Point (3) does not fall in one or other category, as it is not a mental process, but related to the physiology of emotion discussed previously. Under another perspective one can see the four points as the contributions of the different approaches to emotion discussed previously: (1) the phenomenological approach; (2) the cognitive approach; (3) the physiological approach and (4); the behavioural, developmental and social approaches.

Concluding remarks

Emotion always existed, as one form or another, as a part of our existence. However, as the profusion of existing theories sustain, it is difficult to define emotion and distinguish it from non-emotion. Conversely, by its ubiquity, emotion can be approached in many senses in Psychology and other related disciplines. Outside psychology there have been excellent theoretical contributions from biology, neurophysiology, sociology, anthropology, cultural studies, philosophy and history. These theories show how emotions are not only an inter-individual matter, that goes far as language and culture, and have a good heuristic value. Beyond emotion, no other aspect of Human condition is at-ease in adopting a multidisciplinary and plural approach. Any other approach would be too simplistic, and would not make justice to this fascinating field.

Although the phenomenal aspect of emotion is still one of the greatest challenges in the study of emotion, the study of the intentional aspects of emotion has seriously progressed. Historically, decision making was viewed as a cognitive function without input from emotion. However, extensive research on the role of affect during gambling tasks revealed that feelings of emotion experienced prior to making a decision influenced anticipated outcomes and action. Thus, support for cognition as the single factor affecting decision making irrevocably changed. There is a high probability that emotions either facilitate or impair every-day life decisions in terms of anticipating negative and positive outcomes, a fact that became well established in the literature.

3. Symbolic Computational Models of Emotion

The approach taken by many AI researchers and Computer scientists when trying to approach the area of Computational models of emotion has been to seek inspiration in traditional cognitive theories of emotion, extract their more general properties, and create models that computationally simulate those emotional processes.

At the same time, the advent of agent based systems has shifted AI from a more “egocentric” approach of the intelligent systems to an approach where the main focus stands in the link between the environment and such intelligent systems. The area of agents has grown significantly, and agents are seen as entities inhabiting environments that are dynamic, non predictive and non episodic.

Those environments lead agents to perceive, to reason and to act according to its link with the environment. However, this link is also the key feature of some emotion based theories, in particular for appraisal based theories. What distinguishes an appraisal theory from other types of emotion theories is the claim that emotions are elicited by evaluations (appraisals) of events and situations perceived by the organism. Emotions are generated when a particular appraisal is made. Since the perceptual system of the organism is designed to notice change in the environment, it will signal the occurrence of an event that will trigger an appraisal. As a result, emotions are generated.
These appraisal based theories, perhaps due to the close link between perception and emotion, have been the source of inspiration for many of the computational models of emotion found in our research field nowadays. Furthermore, given the nature of some of these theories, where emotions are seen as discrete elements, attaching "symbols" (labels) to those labeled emotions seems an obvious approach to take.

In this review, we will therefore focus on appraisal models for computational systems. We will not try to be exhaustive in the review of several models, but rather present a small set of cases that somehow may have some direct impact onto the goal of achieving anticipation. In fact, several theorists maintain that the appraisal system has evolved to process information that predicts when a particular emotional resource is likely to provide the effective coping for the situation. Thus, emotion plays a crucial the role in the prediction of responses and events.

When building a computational model that includes emotional responses, there are some aspects that need to be considered:

- how is an emotion represented and how an emotion is generated
- how emotion motivates action
- how emotion distorts or guides perception, inference, and learning
- how emotion is communicated to the environment

Furthermore, when considering an appraisal based approach, we need to decide some key elements in the system, in particular:

- what characterizes an emotion
- the appraisal variables (or the appraisal structure) used for the appraisal process
- the processes from perceiving the world to acting on the world

- the emotional state influence on other computational processes such as problem solving and planning.

ALEC

Using a different approach on the role of emotions in hybrid architectures, ALEC (Asynchronous Learning by Emotion and Cognition) architecture (Gadanho, 2003) focuses on emotional and cognitive learning/decision-making capabilities so that agents can adapt to real world environments. Gadanho and Hallam (Gadanho and Hallam, 1998) address the problem of designing an Adaptive Controller for a simulated Robot, operating in a continuous real-time environment. The robot has to make its way in a labyrinth-like environment, delimited by walls and obstacles, searching for valuable energy that it needs in order to survive.

The Robot is constantly consuming energy from its limited-life battery that can only be refilled at specific points available throughout the environment and identified by a detectable light. However, these energy refilling points have themselves limited capability and require a considerable recharge time before they can be used again by the Robot. Therefore, the Robot must constantly search for the best alternative energy points available. The simulation system introduces a certain noise in the values read by the sensors and in the effective power delivered to the motors, creating supplementary problems in both Perception and Control. The Robot has three pre-programmed behaviours available that build its (compound) Action/Behaviour set: Avoid Obstacles, Seek Light and Follow Walls.

Gadanho and Hallam's goal is to develop a controller that is able to optimally choose the appropriate behaviour from the three possible behaviours available, according to the current state of the environment. To solve this problem, they propose an emotion-based architecture (EB architecture) in which a traditional reinforcement learning adaptive system is completed with an emotion system responsible for both reinforcement and behaviour switching. The justification for the use of emotions is that, in nature, emotions are usually associated with either pleasant or unpleasant emotions that can act as reinforcements. The agent has some innate emotions that define its goals, and it learns emotion associations of environment-state and behaviour pairs, which determine its decisions. The agent uses a Q-learning algorithm to learn its behaviour-selection policy while it interacts with its environment.

The ALEC architecture aims at a better learning performance by extending the EB architecture with a cognitive system containing explicit rule knowledge extracted from the agent-environment interaction. The distinct underlying learning mechanisms are consistent with the assumption that, in nature, the cognitive system can make more accurate predictions based on rules while the emotional associations have less explanatory power but can make more extensive predictions and predict further ahead in the future.

ALEC Architecture

The ALEC Architecture is divided into the emotion system and the cognitive system. The emotion system is composed by the goal system and adaptive system of EBII Architecture. The cognitive system is based on the rule system of CLARION (Sun and Peterson, 1998).

Goal System

The goal's system role is to complement a traditional reinforcement-learning adaptive system so that the learning is autonomous. The goal system determines how well the adaptive system is doing. Furthermore, it is also responsible for deciding when behaviour switching should occur.

Goals are explicitly identified and associated with homeostatic variables. These variables are associated with three different states: target, recovery, danger. The state of each variable depends on its continuous value grouped in four qualitative categories: optimal, acceptable, deficient and dangerous. A well being value is derived from the state of the above variables. If a variable is in the target state it has a positive influence on the well-being, otherwise it has a negative influence proportional to the deviation from the target values.

Well-being is also influenced by two specific events: when a homeostatic variable changes from one state to another the well being is influenced positively if the change is towards a more favourable state and negatively otherwise; when some perceptual clue predicts the state change of a homeostatic variable, the influence is similar, but lower in value and dependant on the accuracy of the prediction.

Adaptive System

The adaptive system uses a Q-learning algorithm. Through this algorithm, the agent learns iteratively by trial and error the expected discounted cumulative reinforcement that it will receive after executing an action in response to a world state, called utility value. This utility value can be learned by neural networks with back-propagation, but they have the problem of being overwhelmed by the large quantity of consecutive similar training data and forget the rare relevant experiences. Using an asynchronous triggering mechanism, like emotional state changes (or homeostatic variables state changes) helps resolve this problem by detecting and using only a few relevant examples for training.

Therefore, the state information fed to the neural-networks consist on the homeostatic variables values and other perceptual values retrieved from the robot sensors. At each trigger step, the agent may select performing the behaviour that has proven to be better in the past, or selecting an arbitrary behaviour to improve its information about the utility of that behaviour. The selection function used is based on the Boltzman-Gibbs distribution and consists of selecting behaviour with higher utility value.

Cognitive System

The cognitive system maintains a dynamic collection of rules which allows it to make decisions based on past positive experiences. Each individual rule consists of a condition for activation and a behaviour suggestion. The activation condition is dictated by a set of intervals, one for each dimension of the input space. Since this represents a very large number of possible states, rule learning is limited to cases in which there is a particularly successful behaviour selection.

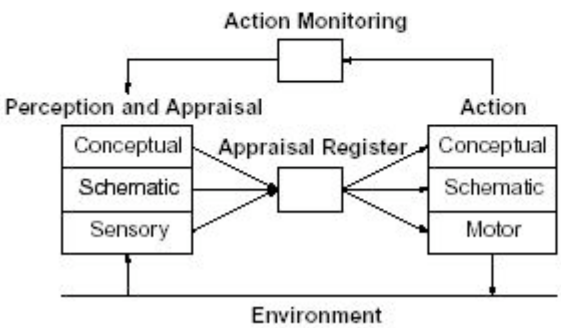
If behaviour is found successful in a particular state, then the agent extracts a rule corresponding to the decision made. Whenever the same decision is made again the agent updates the record of the success rate of the rule. If the rule is often successful, the agent tries to generalize it by making it cover a nearby environmental state. If the rule's success rate is very poor then the agent tries to make it more specific. If this does not improve the success rate, or if the rule covers only one state, then it is deleted. When the cognitive system has a rule that applies to the current environmental state, then the cognitive system influences the behaviour decision, by adding a constant value of 1.0 to the respective Q-value before the stochastic behaviour selection is made.

Experimental results have shown that the agents with both cognitive and emotional systems would not only perform better in their environment, but would also learn faster than agents without such systems.

TABASCO Architecture

TABASCO (Tractable Appraisal-Based Architecture for Situated Cognizers) (Staller and Petta, 1998) presents an agent architecture based on Scherer's theory of emotions. According to Scherer (Scherer, 1984), emotions are a flexible adaptation mechanism that evolved from reflexes and physiological drives. Therefore in TABASCO, emotions are modelled through an adaptive process related with the agent-environment interactions.

Scherer's theory clearly distinguishes three hierarchical appraisal levels: sensory-motor, schematic, and conceptual. The sensory-motor level is based on innate hard-wired feature detectors giving rise to reflex-like reactions. The schematic level is based on schema matching. The conceptual level involves reasoning and inference processes that are abstract, active, and reflective.



Staller and Petta basic idea is that the concept of three level processing, should not only be applied to the appraisal mechanism but should also be used in the action selection process like depicted in the above Figure, TABASCO architecture is composed by four main modules: appraisal and action selection (both of them with three levels), appraisal register, and action monitoring.

Appraisal Component

Like mentioned above, the appraisal component is composed by three layers. The sensory-motor layer consists of a set of feature detectors used to match received stimulus. For instance, it can be used to detect a sudden and intense stimulus or event. This mechanism permits also to quickly determine an approximate value to the valence (good or bad) and intensity of the received event/stimulus.

In the schematic level, the event is matched with schemas (specially social and self schemas) stored in the agent's memory. Schematic processing is fast, automatic, parallel, inflexible, and concrete. It can be thought of in terms of priming and spreading activation. The conceptual level involves abstract reasoning and inference based on propositional knowledge and beliefs. The inferences involved in causal attribution are modelled at this level as well as the evaluation of one's actions in relation to norms or one's self-ideal.

Action Selection

The motor level is responsible for performing lower level actions that can be used by the upper levels (walking, picking objects, etc). One important type of such actions is facial and bodily expressions, which are used to convey emotions to external viewers.
TOBASCO architecture uses the distinction provided by Lazarus (Lazarus, 1991) that states that action tendencies correspond to innate reactions to the environment, triggered by certain stimulus, while coping correspond to a more deliberative and complex process. Thus, the schematic action selection level implements the character's action tendencies by defining rules in the RAP System (Firby, 1989). The conceptual layer is composed by a planner responsible for building more complex sequence of actions.

Appraisal Register

This component mediates between the appraisal and action module. First, it detects and combines the several appraisals performed by the distinct levels. Next, it influences the action selection component by using the combined appraisals. For instance, this module can start the immediate execution of a motor program, activate an action tendency by putting a RAP rule in a state of readiness or even start a long-term planning procedure.
However, it is not necessary to wait for the appraisal processes of all levels to initiate an action selection process. For instance, a startle response occurs immediately after the sensory input has been processed, without waiting for the slower conceptual appraisal.

Action Monitor

Finally, the action monitor component is responsible for the process of "reappraisal". This component keeps monitoring the processes of planning and action execution, sending the results back to the appraisal component in order to generate new appraisals. For instance, if all existing plans and actions are unsuccessfully used to solve a given problem, such information is sent to the appraisal module to generate an appraisal that no solution exists.

Sloman Architecture

Sloman defines a Hybrid Architecture that involves the combination of two other traditional architectures: the "three towers" model and the "three layers" model (Sloman, 2001).
The "three towers" model comprises three parallel subsystems that are conceptually vertical within the Agent Architecture:

- The **Perception Subsystem** that is responsible for extracting data from the environment where the Agent is operating;
- The **Action Subsystem** that allows the Agent to act upon the environment;
- The **Central Processing subsystem** that mediates perception and action and is capable of controlling both of these subsystems.

In order to achieve a globally successful performance, there is usually a great interaction level among these three subsystems. For example, the Action Subsystem may provide direct feedback to the Central Subsystem (proprio-perception) or it may interact with the Perception Subsystem to allow effective coordination of some tasks (e.g.: hand-eye coordination).

The “three layers” model consists of the following layers:

Reactive Layer. This layer includes reactive mechanisms, capable of delivering automatic responses when triggered by specific environment (internal and external) conditions. The reactive layer contains mechanisms capable of implementing pattern matching functions, condition-action rules and other direct input-output functions.

Deliberative Layer. This layer would have resulted from the evolution of some of the mechanisms belonging to the Reactive Layer. The most important evolution in this layer is the development of reasoning capabilities about past, present and future events (what if reasoning). This layer also supersedes the previous one by adding more sophisticated memory systems and symbolic reasoning capabilities.

Meta-Management Layer. This third layer includes mechanisms intended to monitor, evaluate and redirect processes executed in the Deliberative and Reactive Layers. This layer would have been developed in order to provide efficient control of all the machinery operating in the two lower layers. The mechanisms implemented at this layer may be reactive or deliberative.

The Architecture presented by Sloman, superimposes the “three towers” model and the “three layers model” forming a hybrid Architecture with nine distinct conceptual components. The processes running in each of the three layers operate concurrently and deal with Perception, Central Processing and Action at different levels of abstraction.

Motives, Global Alarm Mechanisms and Variable Attention Filter

Sloman observes that Agents operating in complex and real-time environments are usually compelled by several Motives. Some of these Motives will be active simultaneously and can be considered competitors regarding the Agent’s available processing resources. Conversely, other Motives will not be permanently active, requiring special motive generator mechanisms to be activated when certain conditions are found. Since the Agent has limited resource capabilities, some information processing mechanisms (in particular those located in the Architecture’s upper layers) may not be able to respond fast enough to specific conditions occurring in real-time environments (either dangers or opportunities). Moreover, urgent Motives may require processing resources not available at that moment, possibly because they are being used to process less urgent Motives.

“Global Alarm Mechanisms” handle the problem of resource limitation. These Global Alarm Mechanisms receive information from every component in the system and, by using a fast pattern matching procedure, they are able to detect situations whose urgency requires a change in the Agent’s internal processing strategy. When such a situation is detected, Global Alarm Mechanisms immediately send system-wide interrupt signals in order to stop current processes and trigger the whole system’s redirection to conveniently deal with the urgent Motive or situation. This may involve switching the Motive that is receiving processing resources or it may even require stopping computationally heavy processes, followed by the activation of faster reactive mechanisms that can provide a more efficient response to the urgent situation.

Global Alarm Mechanisms can help the Agent achieve a balanced use of its capabilities. However, they may also become very unproductive, if the conclusion of deliberative and meta-management processes is systematically delayed or postponed as a result of frequent interruptions. In order to reduce this possible negative effect of Global Alarm Mechanisms, Sloman has devised the Variable Attention Filter mechanism, which is intended to filter some of the interruptions targeted at specific processes by raising the minimum urgency level that may interrupt the process. This will result in what can be considered as a state of "concentration". Therefore, when Agents need to execute tasks that require continuous deliberative or meta-management processing, the Variable Attention Filter will increase the minimum urgency threshold to ensure that such process will not be interrupted, unless the signal is generated in response to a very urgent condition or Motive.

Emotion

Sloman draws a relationship between Emotion and the interactions established between the Architecture's subsystems. In fact, Sloman argues that Emotional States arise naturally from interactions established between these subsystems, with no need for a dedicated Emotion-generating mechanism. Sloman divides Emotional Phenomena into three categories directly connected to the three layers in the Architecture. Following the definitions proposed by Damásio (Damásio, 1994) closely, Sloman agrees on the existence of Primary and Secondary Emotions but introduces an additional concept: Tertiary Emotions.

According to Sloman, Primary and Secondary Emotions result from the interactions established between Alarm Mechanisms and other subsystems located in the Reactive and Deliberative Layers (Sloman, 1998). Primary Emotions, such as being startled, frozen with terror or sexually aroused (Sloman, 2001), are supported by Alarm Mechanisms located in the Reactive Layer, which are mainly concerned with processing sensory information (inside and outside the environment) and triggering fast Reactive Mechanisms.

Secondary Emotions are supported by mechanisms in the Deliberative Layer and include emotions such as apprehension, relief and other semantically rich emotions that require deliberative capabilities. Secondary Emotions result mainly from Alarm Mechanisms concerned with evaluating internal cognitive responses (e.g.: the chances of success of a risky plan) that are not directly linked to the perceived environment.

Tertiary Emotions, on the other hand, result from the mechanisms located in the Meta-Management Layer. Therefore, Sloman argues that they are probably exclusive to Humans. They include emotions related with thought and attention control such as infatuation, humiliation and thrilled anticipation. These Emotions interfere with Deliberative processes by diverting the attention from current tasks and triggering introspective processes, in spite of the Agent' s attempt to ignore such interruptions.
It is important to note that the perspective followed by Sloman regarding Emotions is focused on explaining their occurrence as a result of Architectural requirements. For the author, the concepts regarding Emotion are a consequence of the evolution of such Architecture. In (Sloman, 1998) Sloman disagrees with the point of view shared by several other researchers (notably Damásio), by stating that Emotions should not be considered an absolute requirement for Intelligence. In fact, Sloman argues that

Emotions are simply side effects of the mechanisms needed to overcome an Agent's resource limitation.

The key concepts around Emotion are Alarm Mechanisms and Interruptions. Emotions are mainly seen as effects of existing control mechanisms intended to: detect situations or motives that need urgent response from the Agent; trigger the appropriate redirection of processing resources at different levels of abstraction.

Émile

Émile system (Gratch, 2000) builds on Clark Elliot's Construal Theory (Elliot, 1992). However, instead of using domain specific rules to determine the appraisal, Émile takes advantage of explicitly storing the agent plans into memory to reason about future possible outcomes and to automatically generate the character's emotional state.

Plans are modelled as a set of STRIPS operators plus a variety of constraints like ordering constraints, binding constraints and planning constraints. STRIPS operators represent the actions that an agent may take in the world and consist on a set of preconditions that must hold in order for the action to be performed and a set of effects that describes how the world changes if the action is performed. Ordering constraints specify that one action must be executed before other action. Binding constraints associate variable names between distinct steps, and protection constraints protect a given effect by assuring that the effect must stay true during some time interval.

Emotion Eliciting Conditions

Construal theory assesses the relationship between events and an agent's disposition through a set of knowledge structures called *construal frames*. These frames specify a set of features called *emotion-eliciting conditions*, which include desirability, expectation status, etc. When an agent perceives an event, he matches it against construal frames to assess its relation to the agent's goals, standards and preferences (is the event desirable? if corresponds to an action, is it praiseworthy?). Next, the emotion eliciting conditions generate and determine the agent's emotions. For instance, if the event is desirable, a *Joy* emotion is created.

Rather than appraising events directly by using domain specific construal frames, Émile adds a level of indirection that significantly generalizes this process. Events influence plans indirectly through the activities of the planner, and Émile appraises the state of plans in memory. Construal frames are created whenever certain syntactic features are recognized in the agent's internal state. For example, whenever the agent adopts a new goal, frames are created to track the status of the goal. Each frame describes the appraised situation in terms of emotion eliciting conditions. These are derived from domain-independent rules that examine the state of plans in memory. Next, the several emotion-eliciting conditions used in Émile are depicted.

Self: This condition specifies whose perspective is being used to form the appraisal. It defines if the appraisal refers to the agent itself or refers to another character. It is possible to reason about the emotions of other agents, by representing their plans and goal.
Desire-Self: This condition describes if the characteristics of the agent's goals are desirables or undesirable (to the agent specified in the previous condition). A goal's local characteristics are desirable if some effect in plan memory establishes the goal and no intervening effect unestablishes it. Otherwise, it is considered undesirable.

Status: The status condition characterizes the expectations of the previous assessment. Are the goal's characteristics confirmed or unconfirmed? A goal frame is desirable but unconfirmed if the goal has an unthreatened establisher (it is established and no other effect possibly undoes it before the goal is needed). The status changes to confirmed once the establishing effect occurs. On the other hand, a goal frame is unconfirmed and undesirable if the goal does not have an establisher or if the establisher is threatened. This state becomes confirmed if the plan probability drops to zero, or if the threat occurs.

Evaluation: This condition assesses if the plan contains a praiseworthy or blameworthy act. This involves reasoning about standards of behaviour. For instance, a goal frame is blameworthy if a protection constraint associated with the goal is threatened by another character's action.

Emotion generation

Given the above emotion-eliciting conditions for a construal frame, Émile generates the following five emotions:

Hope: If the frame is desirable but unconfirmed, a *hope* emotion is created.

Joy: A goal has been achieved. This emotion is generated with a desirable and confirmed frame.

Fear: This emotion is created by an undesirable frame with an unconfirmed status.

Distress: A goal has failed. Detected by an undesirable and confirmed frame.

Anger: Created if the frame contains a blameworthy act.

In order to determine those emotions intensity, Émile relies on goals and plans. Following Neal Reilly approach, Gratch uses a simple model with two intensity variables: probability of goal attainment and goal importance. The probability that a goal will be achieved depends on how one intends to achieve it, i.e., the goal probability is obtained from the plans to achieve the goal. In order to determine such value, one must first supply the probability that an effect of an action will be achieved if the action is executed. Afterwards, it is necessary to determine the probability that an unplanned goal can be successfully achieved.

The next table shows the formulas used to determine the intensity for each of the emotions stated above.

Emotion	Intensity
Hope	Importance(goal) * P(goal)
Joy	Importance(goal)
Fear	Importance(goal) * [1-P(goal)]
Distress	Importance(goal)
Anger	Importane(goal) * P(threat)

The character's emotional state keeps updating continuously. The plan-based appraisal provides a general model of the dynamics of emotion. Whenever plan memory changes, Émile automatically revises its appraisal and probability assessments. For example, an agent may have an important and unestablished goal that leads to an appraisal of fear. But after some effort, the planner discovers a good plan, causing the fear appraisal to retract and a hope appraisal to be asserted.

Émile draws on Velásquez's Chataxis (Velásquez, 1998) model to model the emotional appraisals. Different appraisals act as energy elicitors that excite or inhibit different emotional states, and decay over time. As long as these appraisals persist, Émile decays their intensity by a constant rate. The decaying intensities are put into different buckets according to their emotional label. Thus, if there are several appraisals of "fear", they are added together to an overall fear intensity. The activation of a given emotion can be excited by other emotional states (e.g., joy excites hope) and inhibited by others.

Behaviour

The behavioural component is built upon STEVE (Rickel and Johnson, 1998). STEVE plays the role of the planner/executer that maintains a representation of the world state, and develops, executes and repairs plans that achieve the agent's goals. The Émile system uses the character's emotional state to select gestures, facial expressions and to alter speech generation.

Although Émile focuses mostly on plan-based appraisal, Gratch proposes a series of interesting mechanisms (in addition to the obvious ones described above) in order to use the character's emotional state to influence the reasoning process. As suggested by Damásio, the emotional state can act as search control, focusing cognitive resources on specific goal or threats. For instance, one could focus planning on the portions of the plan generating the most intense appraisals. Nevertheless, it can also be used to alter the overall character of problem solving. For example, according to Sloman, negative emotions seem to lead to narrow focused problem solving while positive emotions lead to broader problem solving, which attempts to achieve multiple goals simultaneously.

Carmen's Bright IDEAS

Carmen's Bright IDEAS (CBI) (Marsella, Johnson and LaBore, 2000) presents an interactive health application designed to improve the problem solving skills of mothers of pediatric cancer patients. A mother learns by making decisions on behalf of a character in a virtual environment, and seeing the consequences of her decisions. The controlled character, Carmen, mirrors the mother's own problems. Carmen has a nine-year-old son with pediatric leukemia and a six-year-old daughter.

CBI is a three act interactive drama. In the first act, the mother is presented with a sequence of situations, dramatizing Carmen's problems. These provide back-character story and help encourage the mother to empathize with Carmen. In the second act, Carmen discusses her problems with a clinical counsellor, who suggests she use a problem solving technique called Bright IDEAS to help her find solutions: **I**dentity a solvable problem, **D**evelop possible solutions, **E**valuate your options, **A**ct on your plan and **S**ee if it worked.

The human mother interacts with the drama by making choices for Carmen, such as what problem to work on and how she could cope with the stresses she is facing. The learner can select alternative internal thoughts for Carmen (presented as thought balloons).

The last act is presented as a linear sequence of scenes, where the scenes depend upon the decisions that Carmen made earlier in order to solve her problems.

Agent Model

The agent model (used to build each onscreen character) is composed by several modules. The problem solving module corresponds to the agent's cognitive layer (goals, planning and deliberative reactions). The dialog module determines how to use dialog to achieve goals. The emotional appraisal module is responsible for emotionally evaluate (appraising) events. Finally, the behaviour generation module constructs the agent's behaviour and passes that behaviour to an animation program.

In CBI, the world-events that the agents process and appraise are dialog annotations. The annotations, along with the problem-solving context, reveal the meaning of what is being said. Agent process dialog in the following order:

"Hear" Dialog Line if there is one.
Appraise dialog and pass result to Behaviour Generation.
Decision-making - form intent to perform a dialog act.
If agent wants to speak goto 5 else goto 1.
Appraise step 3 decision-making and pass result to Behaviour Generation.
Compute dialog pause based on emotional state and pass to Behaviour Generation.
Pass phrase and annotations to Behaviour Generation to build the behaviour program that will be executed in parallel with this phrase.
Appraise phrase just spoken and pass to behaviour generation.
If phrases remain go to 6, else go to 1.

Appraisal at (2) starts the emotional appraisal that is used to determine the character's emotional state. Decision-making (3) comes with a plan to say something. The second appraisal (5) sets emotional state and expressions based on the previous decision. The behaviour program built at (7) creates a parallel and sequential structure that includes gestures, facial expressions, etc (dependent on the emotional state). The final appraisal (8) sets emotional state in reaction to content transmitted in each phrase.

An emotional model of appraisal and coping

CBI presents an innovative emotional model based on the view of Lazarus (Lazarus, 1991). His work organizes human behaviour around appraisal and coping. Appraisal leads to emotions by assessing the person-environment relationship. For instance, is an event relevant to the agent's motivations and is it congruent or incongruent to those motivations? More concretely, Marsella modelled this assessment around the concept of ego identity: how an event impacts an individual's ego-identity? Ego-identity is the individual's collection of concerns for self and social-esteem, social roles, moral values, self-ideals as well as concerns for other people's well being.

In CBI, ego-identity is modelled as a collection of role ideals (Carmen wants to be a good mother), concerns (good-mothers want their children to be healthy) and responsibilities (good-mothers are responsible for their child's behaviour). In order to generate appraisals, CBI uses a series of appraisal rules that refer to these representations. For instance, it is Carmen's concern for her son's well being that induces sadness. Such appraisal is attained by the two following appraisal rules: *If an event violates a concern, it is negative; if talking about negative event then increase anger*.

Coping is the process of dealing with emotions, either by acting externally on the world (problem-focused strategies), or by acting internally to change beliefs or attention (emotion-focused strategies) in order to mitigate negative emotions. For example, a problem-focused way to attempt to deal with a loved one's illness, is to take action that gets them medical attention. Emotion focused strategies may include avoiding thinking about it, focusing on the positive or denying the seriousness of an event.

In order to address coping, CBI performs a secondary appraisal to evaluate four factors: *accountability, expectancy, problem-directed coping potential* and *emotion-directed coping potential*. Accountability establishes who is to blame for a motivationally incongruent event. Expectancy establishes whether there is hope that things will get better. Coping potential are an assessment of how effectively the agent will be able to cope. The two appraisals establish the character's emotional state that is used to control the behaviour and coping strategies used by the characters. In addition to annotated dialogs, problem-solving activities can also generate appraisals. For instance, successful problem solving, increases confidence in problem directed coping, which in turn impact Carmen's willingness to engage in problem solving.

Coping is key to the agent's selection of dialog and its response to it. Carmen may choose an evasive coping strategy and select dialog consistent with that strategy using the coping annotations. For example, the Carmen agent's emotion model appraises the discussion of Diana's (Carmen's child) tantrums as a source of distress because of her concern for Diana and because failure to control Diana may reflect on her ability as a mother. Her response to this stress may be to blame Diana and trivialize her tantrums by saying she is just being babyish.

Interaction Model

The interaction between Carmen, Gina (the counsellor) and the mother is crucial to the drama in CBI. In the interaction model used, called rubber-band model, both Gina and the learner exert influence over Carmen but the influence is partial and mediated by Carmen's own cognitive and emotional dynamics. It is Gina's job to keep the social problem solving on track so that the story proceeds to a successful outcome by effectively responding to Carmen's cognitive and emotional state, at times motivating her through dialog to work through the steps of IDEAS on some problem or alternatively calming or reassuring her.

Gina reassures or sympathizes when Carmen is distraught but prompts Carmen to address the current step in the current dialog strategy when Carmen is less distraught. If Carmen's emotion model leads her to respond inappropriately, Gina has to decide how to repair this failure by directing Carmen towards emotion-directed or problem-directed coping by giving either emotional or instrumental support.

Both Gina dialog, and the user choices (by selecting Carmen thought's such as the problem to work on and how se could cope with it) influence the cognitive and emotional state of the agent playing Carmen, which in turn impacts her behaviour and dialog. The cognitive and emotional dynamics within the Carmen agent ensures that Carmen's behaviour is believable at all times, regardless of how Gina and the learner may be influencing her.

Physical Focus

The character's emotional state is used to change his facial expression, gestures and body posture. To address this concern, the agent architecture relies on a Physical Focus model that bases an agent's physical behaviour in terms of what the character attends to, how they relate to themselves and the world around them.

CBI models four distinct focus modes: strong body-focus, body-focus, transitional and communicative. Strong body focus is characterized by a self-focused attention, away from the conversation and problem-solving behaviour. It is usually associated with considerable depression or guilt. The agent exhibits minimal communicative gestures such as deictic or beat gestures, presents very paused or inhibited verbal activity and tends to avert the gaze and to perform self-punitive hand to body stimulation (e.g. squeezing the forearm). Body-focus mode is similar to the previous mode but shows a more moderate withdrawal.

Transitional indicates less depression, some willingness to take part in the conversation and to perform problem solving, and a closer relation to the listener. Physically, it is characterized by hand to hand gestures, hand to object gesture and with some muted or stilted communicative gestures.

Communicative mode indicates a full willingness to engage in the dialog and problem solving. It is marked by the agent's full range of communicative gestures.

Rules map the current aggregate emotional state into a specific focus mode. Higher levels of guilt or sadness induces transitions toward body focus, while higher levels of hope or anger induces transition towards communicative. The transitional focus lies between these extremes.

Mission Rehearsal Exercise

The Mission Rehearsal Exercise (MRE) system was designed to create learning environments where one can experience high-stress social situations in the relative safety of virtual reality. The system aims to teach decision-making skills in such highly evocative situations. Intelligent agents control characters in the virtual environment with which the participants must interact in the course of their training.

Such virtual humans must incorporate emotional models that can respond in reasonable ways to whatever circumstances the user is allowed to create. They must identify plausible emotions to express, and model the typical coping strategies people use in emotional situations. In order to achieve such results, MRE presents an integration of two research efforts focused in creating engaging and believable characters. Gratch's Émile system focuses on how emotions arise from an evaluation of events regarding the agent's goals and plans. Marsella's CBI (Marsella, Johnson and LaBore, 2000) system addresses the complementary aspects of how emotions impact on the agent's behaviour.

Appraisal

The appraisal process is based on Émile as described above. However, some improvements were made. Now Émile models conditional plans: plans may contain sensing actions with indeterminate effects, and conditional plans can be construct to cover alternative possible outcomes of sensing actions. The previous model only tracked the utility of goals (positive utility) and ignored possibility that an action could have an effect that was undesirable. The current model allows the effects of actions to have positive or negative utility, which allows assessing the impact of any non-goal related side-effects of actions. Finally, instead of using just two appraisal variables (goal probability and goal importance), the new version models the following appraisal variables (merged from CBI and Émile):

Goal relevance: are the consequences of an event relevant to an organism's goals.
Desirability: how desirable are the consequences.
Likelihood: how likely are the consequences.
Causal attribution: who is the causal agent underlying the event and do they deserve credit or blame.
Coping potential: a measure of an agent's ability to reverse negative or maintain positive circumstances.

It is important to point out that MRE appraises each consequence of an event separately. For example, if an event has positive and negative effects such as getting a bad tooth removed, the event generates strong negative and positive appraisals. This ability to separately consider different aspects of the same event plays an important role in certain coping strategies, which attempt to focus on one aspect to the exclusion of the other.

Focus

The character's emotional state maintains numerous simultaneous appraisals that are updated by any change to the causal interpretation. The term causal interpretation refers to the configuration of beliefs, desires, plans and intentions that represents the agent's current view of the agent-environment relationship. Having such high number of simultaneous appraisals raises the issue of what focuses the virtual humans on particular emotions that need to be coped with.

The agent possesses a number of cognitive operators that access or alter the causal interpretation. Such operators include planning related operators (update a belief, update an intention, etc), dialogue related operators (understand speech, output speech, etc), and execution operators (monitor an effect, action initiation, etc). When a cognitive operator accesses any part of the causal interpretation, any appraisal frames associated with that portion of the data structure are brought into focus.

Coping

Emotions do not serve just to modulate facial expressions or physical focus modes. They are also power motivators. The coping model proposed by Marsella in CBI was further developed and put to use in MRE.

Selecting a coping strategy is a four-stage process: (1) identify a coping opportunity, (2) propose alternative coping strategies, (3) assess coping potential and (4) select a strategy to apply.

Identifying coping opportunities

When a cognitive operation is performed, the coping process identifies any associated appraisals that could motivate coping. As such, the coping process creates a coping elicitation frame that consists on the following fields:

Focus Agency: The agent or object that triggered the cognitive operation.

Interpretation-objects: Any events or states referenced in the cognitive operation. For each interpretation object, the coping process identifies the strongest positive and negative appraisal associated with the object.

Agency-max: Corresponds to the max emotion that the agent believes that the subject referred in focus-agency has about the same referent.

Max-interpretation: Interpretation object with the highest appraisal.

Potential Responsibility: Potentially responsible parties for the interpretation.

If the intensity of the strongest appraisal of the max-interpretation surpasses some pre-specified threshold, the coping elicitation frame is considered a coping opportunity.

Propose alternative coping strategies

For each coping opportunity, several alternative proposals are made regarding the coping elicitation frame properties. Each coping strategy is composed by a set of preconditions that define its applicability. The following table shows the several implemented coping strategies.

Strategy	Conditions	Effects
Planning	Possible future event has desirable effect (facilitates desired state or inhibits undesired state)	Assert intention that event occur
Acceptance	An intended future state (goal) seems unachievable (e.g., no viable plan exists)	Retract intention
Positive reinterpretation	Past event or intended future event with undesirable effect has desirable side-effect	Increase intrinsic utility of desired goal
Mental disengagement	Desired goal seems unachievable	Decrease intrinsic utility of desired goal
Denial/Whishfull thinking	Effect of past event or intended future event has undesirable effect	Decrease probability of undesirable effect
Shift/Accept blame	Event has undesirable/desirable effect and ambiguous causal	Assert blame/credit to one of the

attribution ambiguous causal agents

The planning strategy consists in forming an intention to take an action whose effect achieves the desired state or blocks threats to the desired state. If the max-appraisal is positive, the strategy asserts a preference to maintain this state. Otherwise, the strategy identifies actions that would overturn the threatening circumstances.

Acceptance is the recognition that a negative appraisal is unavoidable. If the maximum appraisal is a threat to a desirable intended state, this strategy proposes dropping the intention. While the threat is still appraised as undesirable, it should come less often into focus, since planning operations will no longer reference the state.

Positive reinterpretation consists in finding some direct or indirect consequences of the event and emphasizing it by increasing its utility for the agent. This strategy will lead negative events to be reappraised in a more positive way.

If a previously identified state seems unachievable, mental disengagement can be applied by reducing the intrinsic utility of the state. In this manner, the emotional charge associated with the event is lowered.

Denial works by denying the reality of an event. If the most intense appraisal associated with the coping frame is negative, the denial strategy is proposed to lower the probability of the effect generating the negative appraisal. As the probability lowers, so does the intensity of the negative appraisal.

Finally, shift blame can be used to manipulate blame. This strategy can be applied when an event has an undesirable effect and a potentially ambiguous causal attribution. The strategy identifies possible blameful individuals and shifts blame to them.

Assess coping potential

In order to determine the coping potential of a given strategy, the coping process maps the strategy effects to the causal interpretation and checks if it would alter the initial appraisals in a desired way. A greater change to a more desirable appraisal, leads to a greater coping potential.

The agent’s personality is used to influence the final result of the coping potential for each strategy, making it higher or lower according to the character personality and the situation. For instance, a pessimistic character with low self-esteem would prefer to use emotion-focused coping than to act on the environment.

Select one strategy

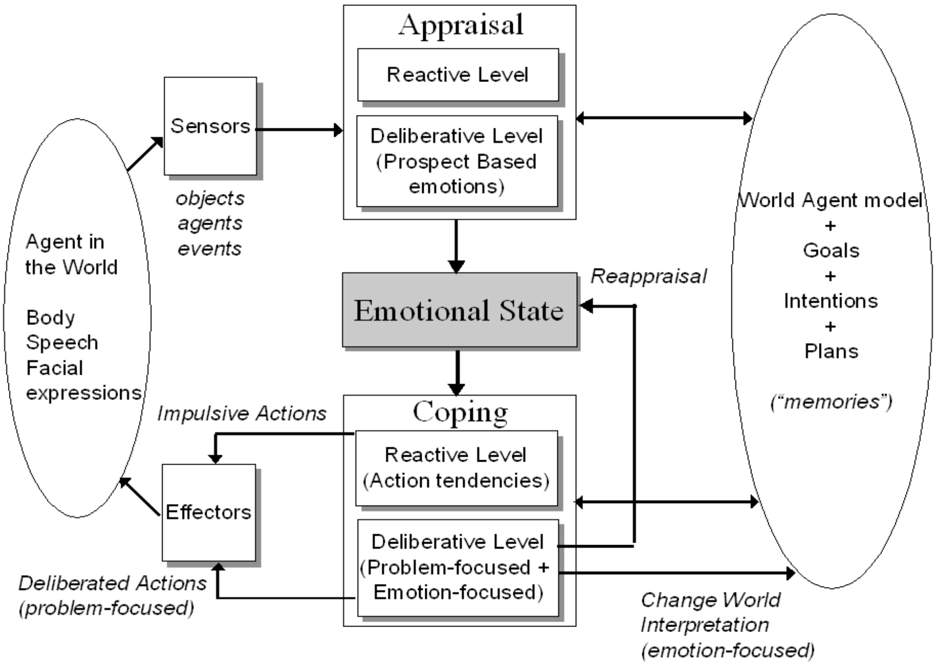
Once the coping potential has been assessed, the strategy with greatest potential is applied. However, in addition to operating in isolation, coping strategies may work in tandem, as long as the various strategies do not conflict in their manipulation of the causal interpretations. Remaining ties are resolved arbitrarily.

Fear-Not

The architecture used in FearNot! (Dias, 2005) presents two levels for action selection. The cognitive layer is responsible for the agent’s plan behaviour, while the reactive layer implements the character’s action tendencies. Action tendencies represent hardwired reactions to emotions and events that must be rapidly triggered and performed. Thus, the character must be able to react to an event and perform an action tendency almost immediately.

Since the action tendencies depend on the character’s emotional state, triggering such actions can only be made after the appraisal process. However, the cognitive appraisal depends on the agent’s plans and can take some time: when an event is received, the

planner has to update all active plans (according to the event) even before the generation of emotions starts. For that reason, we applied the same two-level distinction to the appraisal process.
While the deliberative level generates prospect-based emotions (Hope, Fear, Satisfaction, etc) based on the agent's plans and goals, the reactive level generates all other types of Ortony, Clore and Collins (OCC) emotions (fortune of others, well being, attribution, attraction) using a set of domain dependent emotional reaction rules as used by Martinho in S3A (Martinho, 1999).



Appraisal

Reactive Appraisal

The reactive appraisal process is based on a set of emotional reaction rules as proposed by Martinho (Martinho, 1999). These rules, which are based in Elliot's Construal Theory (Elliot, 1992) generate the majority of OCC emotion types: Well Being emotions, Attraction emotions, Fortune of Others emotions and Attribution emotions. An emotional reaction rule is composed by a domain specific construal frame extended with values for some of OCC emotion intensity variables. The following figure shows three examples of these rules.

Reaction Rule	Reaction Rule	Reaction Rule
Event	**Event**	**Event**
Subject: --	Subject: SELF	Subject: --
Action: Cry	Action: Look-At	Action: Push
Target: --	Target: Book	Target: Book
Parameters: --	Parameters: --	Parameters: --
Appraisal Variables	**Appraisal Variables**	**Appraisal Variables**
Desirability: 9	Desirability: --	Desirability: 5
DesirabilityForOther: -10	DesirabilityForOther: --	DesirabilityForOther:--
Praiseworthiness: -5	Praiseworthiness: --	Praiseworthiness: --
Like: --	Like: -5	Like: --

All events have four fields. The Subject property specifies who made the action, the Action field corresponds to the action name, the Target specifies toward whom or what the action is directed, and finally the last field, parameter is used when the action needs additional parameters.

When specifying an emotional reaction rule, one can leave any field of the event unspecified. If so, any event matches the unspecified field, for instance the rightmost reaction rule matches any event that corresponds to a *Push Book* action, disregarding whoever made the action.

Although several rules may match against a given event, only the most specific one is selected (has described bellow), and its corresponding appraisal variables are used to determine and generate the character's emotion.

Cognitive Appraisal

The deliberative appraisal starts by monitoring the execution of selected actions. The deliberative component associates every action with a set of effects which can occur with a given probability. So, when an action finishes, the deliberative layer checks if all predicted effects occurred or not and updates their probability accordingly. Additionally, it updates all existing plans according to the action result.

Instead of writing domain specific reaction rules to handle prospect based reactions, a similar approach to the one used in the Émile System (Gratch, 2000) is followed to take advantage of explicitly storing the agent plans state and intentions into memory. The agent's deliberative goals and plans are used to generate prospect based emotions. Our model uses two of OCC goal types, *active-pursuit* goals and *interest goals*. Active-pursuit goals are goals that the characters actively try to achieve, like going to a dentist appointment. Interest goals represent goals that a character has but does not pursue, as for instance wanting his favourite team to win a match, or avoiding getting hurt.

Active-pursuit goals are characterized by a set of activation and success/failure conditions. Every time the agent receives a new perception it checks all deactivated goals to determine if any of them has become active. If so, an intention to achieve the goal is added to the intention structure. Initial hope and fear emotions based on the goal importance are created in this process. Afterwards, the deliberative layer must choose between the existing intentions to continue deliberation.

Attention and Anticipation

Emotions can be used to select the most relevant intention. According to Sloman (Sloman, 2001), emotions are an efficient control mechanism used to detect situations or motives that need urgent response from the agent, and to trigger the appropriate

redirection of processing resources. The idea is that the intentions generating the strongest emotions are the ones that require the most attention from the agent, and thus are the ones selected by the planner to continue deliberation. Just like if the character's thoughts are guided by the most intense feelings.
Using emotions to choose the most important intention is crucial to achieve a differentiated reasoning in characters with distinct personalities. Emotions are biased by the character's personality. For example, a fearful character experiences Fear more easily and thus the dominant emotion is frequently Fear. Therefore, the character thoughts are usually driven by fear which makes him give more attention to goals that seem unachievable, and making him give up goals that threaten other interest goals much more easily.
When a plan (the best plan built so far for an intention) is brought into attention by the reasoning process, it generates and updates prospect based emotions. This process of generating emotions is similar to the notion of focus used in MRE System (Gratch and Marsella, 2003). The difference is that the focus is not over a specific operator but is made over the entire plan. When a plan is brought into focus, it generates the following prospect based emotions:
Hope: Hope to achieve the intention. The emotion intensity is determined from the goal's importance of success and the plan's probability of success.
Fear: Fear for not being able to achieve the intention. The emotion intensity is determined from the goal's importance of failure and the plan's probability of failing.
Fear: Fear for not being able to preserve an interest goal. This emotion is generated if the plan contains any inter-goal threat.

Interest goals may specify protection constraints. These allow the modelling of conditions that the character wishes to protect/maintain. Whenever an action is added to a plan, a conflict between the action's effects and existing protected conditions may arise. This conflict is named an inter-goal threat. When the best plan is brought into focus, if it has any inter-goal threat, in addition to the normal emotions, it also generates a fear emotion according to the respective Interest Goal that is being threatened. This emotion's intensity depends on the likelihood of the threat succeeding and on the interest goal's importance.
The attention process is also tightly connected with anticipation since the emotions triggered are the result of an anticipated future situation (thus such emotions are named prospect based emotions). In order to illustrate our point, consider the following example: John, a fearfull 9-year old character has the goal of fighting back when he is insulted or bullied by other character. In addition he also has the interest goal of not getting hurt. Suppose that John is insulted by another kid in school (Luke, a usual bully), in that situation the deliberative layer will assert the intention to fight back and will develop a plan to achieve such goal. However all the actions that John considers to fight back have some likelihood of getting hit back (the bully is stronger than him). So, when such plans are brought into attention, a threat to John's interest goal of not getting hurt is detected and thus John feels fear for his anticipation of getting hit by the victim.

Action Selection/Coping

As soon as the appraisal process is finished, the corresponding layer can start its action selection/coping mechanism. The reactive layer uses the generated emotions to activate and select between action tendencies, while the deliberative layer tries to build and execute plans consisting on a set of actions to achieve active goals.

Reactive Layer

The reactive level implements the character's action tendencies. It consists on a set of action rules that are available according to the character's emotional state. In the action phase, the reactive layer starts by determining which actions can be executed. This is done by testing the action *Preconditions*, if all of them are true, then the action can be executed. The resulting action set is matched for activation against all emotions in the character's emotional state (every rule specifying an emotion that triggers the action). In the last stage of this process, the set of rules with positive matches is used to select the action that will be executed. The action rule triggered by the most intense emotion is selected for execution. If more than one action rule is selected (triggered by the same emotion), the most specific one is preferred.

Deliberative Coping

The coping strategies performed over the selected plan depends on the character's emotional state and personality. Inspired by MRE and CBI (Marsella, Johnson, and Gratch, 2000), the proposed model uses two types of coping: problem focused coping and emotional focused coping. Problem focused coping focus on acting on the environment to cope with the situation, thus it consists on planning a set of actions that achieve the pretended final result and executing them. Emotion focused coping works by changing the agent's interpretation of circumstances (importance of goals, effect's probability), thus lowering strong negative emotions. This is often used by people, especially when problem focused coping has low changes of success.

Coping strategies present an activation model simpler than the one used in MRE. When the planner analyses the plan, it tests the several strategies activation conditions (with a specific order) and applies every one that satisfies its conditions. Next table presents the several coping strategies and its activation conditions.

Activation Condition	Strategy	Effect
Plan probability very low	Acceptance	Drop the plan
Inter-goal threat detected, current goal's emotion stronger than interest goal's emotion	Acceptance, or Denial/Whishfull thinking	Accept the failure of the interest goal (ignore the threat) or lower the threat's probability
Inter-goal threat detected, interest goal's emotion stronger than goal's emotion	Acceptance	Drop the plan
Acceptance strategy applied	Mental Disengagement	Lower the goal's importance
Causal Conflict detected	Planning or Denial/Whishfull thinking	Use promotion, demotion, or lower the conflict probability
Open Precondition	Planning	Add a step that achieves the precondition
Consistent plan	Execution	Execute an action

without open preconditions

Acceptance is the recognition that something is not possible to achieve or protect/maintain. If the selected plan's probability is lower than a given threshold, the character thinks that it's not worth the time to try to improve the plan, since adding more actions will not increase its probability of success, and drops the plan. The threshold used depends on the character's mood. For instance, if the character is in a good mood he gives up plans less easily.

Whenever an acceptance strategy is applied, mental disengagement is also applied. Mental disengagement works by reducing the goal's importance. Since acceptance will frequently lead to goal failure, lowering the goal's importance reduces the intensity of the negative emotions triggered when the goal fails.

When the planner detects an inter-goal threat in the plan, it activates more coping strategies. If the threathened condition generates stronger emotions than the goal's emotions, the current plan is dropped. In the opposite situation, the character can either accept the interest goal's failure (by removing the protected condition) or use wishful thinking to cope with the fear emotion. Wishful thinking works by denying the reality of an event or by thinking that something bad will not happen. This strategy lowers the threat probability by lowering the probability of the effect that threatens the condition.

Finally, when the planner achieves a consistent plan with no open preconditions it has reached a solution. This solution that corresponds to a partial ordered plan is then executed by repeatedly choosing and performing any of the next possible actions.

It is important to point out that since part of the coping strategies are triggered by emotions and mood, the emotional state and personality influence the strategies applied and hence the overall reasoning performed by the characters. For instance, a fearful character has much more chances to drop an active pursuit goal if it presents threats to other goals.

4. Anticipation and Cognition

At the highest level, it is clear that a prominent if not overwhelming part of our own everyday behaviour is based on the tacit employment of predictive models. Similar examples to the bear cited in the introduction can be multiplied without end, and may seem fairly trivial. But similar anticipatory behaviour can also be found at lower levels, where there is no question of learning or of consciousness: the profusion of anticipatory behaviour at all levels of biological organization is an example of that.
In his seminal work regarding anticipatory systems, Robert Rosen (Rosen, 1985) stated that "the failure to recognize and understand the nature of anticipatory behaviour (…) is the necessary consequence of the entire thrust of theoretical science since earliest times not arbitrary, but obey definite laws which can be discovered". By that, Rosen exposed a recurring basic pattern of causality and laws, arising initially in physics and generalized over the years stating that: "in any law governing a natural system, it is forbidden, to allow present changes of state to depend upon future features". Past state, perhaps, in systems with memory; present state certainly; but *never* future states. A denial of causality thus appears as an attack on the ultimate basis on which science itself rests.
However, if we consider the behaviour of a system which contains a predictive model, and which can utilize the predictions of its model to modify its present behaviour, if we further suppose that the model can approximate by its predictions the futures events with a high degree of accuracy, then this system will behave as if it was a true anticipatory system, i.e. a system which behaviour depends on future states. This system will not violate our notions of causality, but since we explicitly forbid present changes of states to depend on future states, we will be driven to understand the behaviour of such system in a purely reactive mode; i.e. one in which present change of state depends only on present and past states.

Universality versus parsimony

The reactive paradigm is universal. Given any system behaviour which can be described sufficiently accurately, regardless of the manner in which it is generated, there is a purely reactive system exhibiting precisely this behaviour. Any system can be simulated by a purely reactive system.
It might appear that this universality makes the paradigm adequate for all scientific explanations, but this is not always the case. For instance, the Ptolomaic epicycles are also universal, in the sense that any planetary trajectory can be represented in terms of a sufficiently extensive family of them. The reason that the Copernican scheme was considered superior to the Ptolomaic epicycles lies not in the existence of trajectories that can be represented by the epicycles, but arises from considerations of parsimony, as embodied for instance in Occam's Razor. The universality of the epicycles is regarded as an extraneous mathematical artefact irrelevant to the underlying physical situation, and it is for this reason that a representation of trajectories in terms of them can only be regarded as a *simulation* and not as an *explanation*.
Robert Rosen believes that it is precisely the universality of the reactive paradigm which has played the crucial role in concealing the inadequacy of the paradigm for dealing with anticipatory systems.

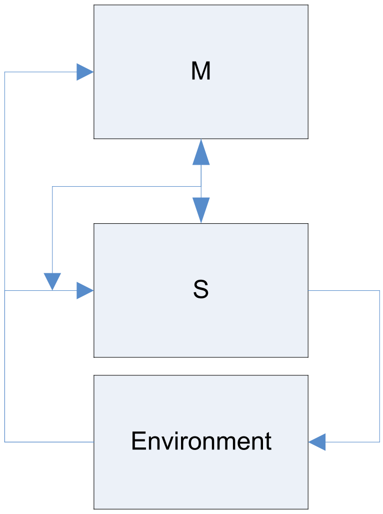
ANTICIPATORY SYSTEM

Let us suppose that we are given a system S, which shall be the system of interest, and which we will call the object system. For simplicity, let us consider that S is a non-anticipatory dynamic system.

With S we shall associate another dynamical system M, which is a model of S. We require that S is parameterized in real time, and that M is parameterized by a time variable that goes quicker than real time. In this way, the behaviour of M predicts the behaviour of S; by looking at the state of M at time T, we get information about the state that S will be in at some time later than T.

We shall now allow M and S to interact with each other. We shall suppose that the system M is equipped with a set of effectors E, which allow it to operate either on S itself, or on the environmental inputs of S, and change the dynamical properties of S.

The following figure represents such a system.



If we put this system into a single box, that box will appear to us to be an adaptive system in which prospective behaviours determine present changes of state. We will call his system an anticipatory system[2].

Anticipation and Emotions

M predicts the future of S, because its trajectories are parameterized faster than those of S. But how is the predictive information used to modify the properties of S through the effectors system E? Let us imagine the state space of S (therefore of M) to be partitioned into regions corresponding to "desirable" and "undesirable" states. As long as the trajectory in M remains in a "desirable" region, no action is taken by M through the effectors E. As soon as the trajectory of M enter the "undesirable" region, which correspond to the expectation of the trajectory of S moving into this "undesirable" region at a later time, the effectors system is activated to change the dynamics of S in such a way to keep the trajectory of S in the "desirable" region.

This is strikingly similar to what happens with emotions! Emotions can feel either good

[2] Rosen called it quasi-anticipatory as M is not a perfect model of S.

or bad and we can think of emotions as being at least partly and usefully classified by where they lie along a good-bad, or positive-negative dimension (generally refereed as the *valence* of the emotion). We can capitalize on what we know of emotions to allow for a more systematic approach to the definition of desirable and undesirable regions.
Rosen pointed to several problems when designing an anticipatory system: the definition of the "steering" variables in S or in the environment of S, through which the dynamical properties of S can be modified, the importance of choosing the right effectors system E and, the problem of programming such a system, a task that ultimately depends on the character of the regions we consider "desirable" and "undesirable". Affective computing may give us some useful insight on how to create such a system.

Side Effects

One important issue when designing anticipatory systems is side effects, a term describing unavoidable and usually unfortunate consequences of employing "therapeutic" agents (as our effectors system E) and that result in unplanned and unforeseeable consequences on system behaviour arising from the implementation of controls designed to accomplish other purposes.
Side effects generally arise, even if the model of the system is perfect and the effectors perfectly designed and programmed, because of inherent system properties. Remember that S is a real system, whereas M is simply a model of a particular functional activity of S. Hence, there are many degrees of freedom of S that are not modelled by M, through which non-modelled interactions can take place. As such, the effectors system E will generally have other effects on an object system S than those planned, which may change the planned interaction between M and S. Side effects are unavoidable consequences of the general properties of the systems and their interaction: they are unpredictable and inherent to the process no matter how well the process is carried out.
Rosen shows that it is impossible to remove all side-effects by augmenting the underlying model, or by controlling each side-effect separately as they appear, as this strategy faces an incipient infinite regress, similar to that pointed by Gödel in his demonstration of the existence of improvable propositions within any consistent and sufficiently rich system of axioms.
An important question is then: how can we update and improve the model system M, and the effectors system E, on the basis of information about the behaviour of S itself?
Although not mentioning Emotions, Rosen also remarked that, as the defect of any part of a sensory mechanism leads to a particular array of symptoms, it should be possible to develop a definite diagnostic procedure to trouble-shoot a system of this kind, by mimicking the procedures used in neurology and psychology: "indeed, it is amusing to think that such (…) systems are capable of exhibiting syndromes (…) very much like (and indeed analogous to) those manifested by individual organisms".
One such an approach will be described later in the document.

5. Anticipatory Affective Systems

WP5 (Emotion and Anticipation) main goal in Mind Races is to identify how emotions and anticipating states relate to each other, provide with models depicting the integration of both concepts, and compare these models both theoretically and in action.

As a first step towards this goal, a set of scenarios was presented in two discussion cycles, in the attempt to identify potentialities and limitations of current anticipatory affective systems, both symbolic, and sub-symbolic. The architectures presented at the kick-off meeting by the partners were further developed and discussed in some detail from the point of view of integrating emotions into the proposed scenarios.

Although it was impossible to reach a common agreement on what an emotion is, a concordance was reached regarding essential relations between emotion and anticipation, in the attempt to separate the structural and psychological relations between the two concepts. Although the relationships are intertwined, an important step was made towards the definition of what an anticipatory affective system is when compared to "standard" anticipatory systems.

Based on the scenarios presented during the two discussion cycles, the mechanisms characterizing an anticipatory affective system were identified and a preliminary blueprint for an integration framework strengthening our results was laid, based around the concept of "anticipatory continuum".

Affective Anticipatory vs Anticipatory

WP5 quickly realized that it would be virtually impossible to reach a common agreement on what an emotion is. Consider, for instance, surprise, a central concept in anticipatory systems, as non-anticipatory systems cannot be surprised. Not everyone considers surprise to be a 'really' affective state or emotion, even if all concur on that it should be modelled in cognitive systems dealing with the future. Other affective states, as relief or disappointment, were perceived otherwise.

Although the term emotion changes according to the architecture supporting it, a concordance was reached regarding the possible relations between emotion and anticipation, an attempt to disentangle very different structural and psychological relationships between the two concepts. Broadly, three different relationships between emotion and anticipation were distinguished:

Emotions eliciting an anticipatory behaviour (the tendency to act);

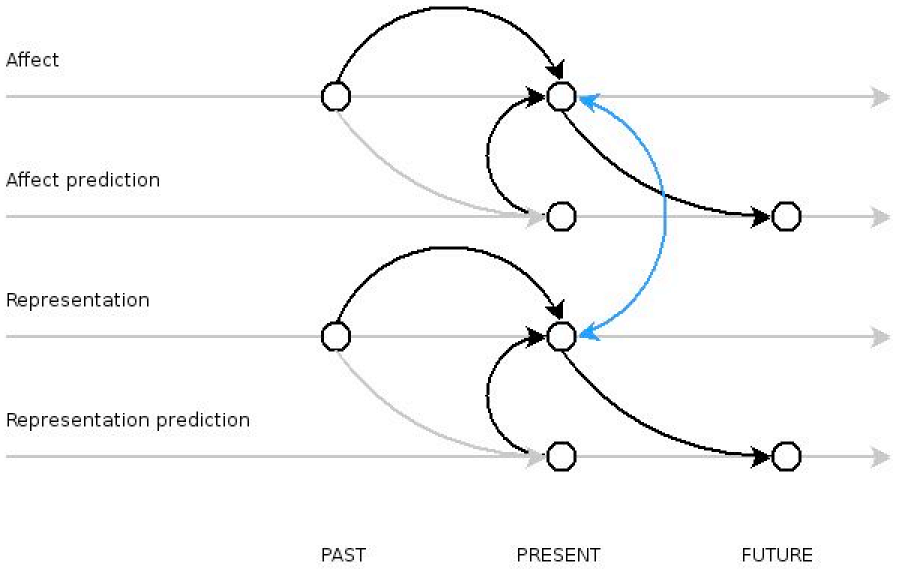
Emotions resulting from eliciting/confronting anticipatory representations (where emotions can be seen as such a representation);

Anticipating future emotions (in the sense that there is a special pathway for emotions when compared to other representation pathways);

Although the relationships are intertwined, an important step was made towards the definition of what an anticipatory affective system is when compared to "normal" anticipatory systems, as not all Mind Races architectures should include affective/emotional processes. The different scenarios proposed during the discussion cycles aimed at investigating all theses different mechanisms and accounted for their possible interactions. As a result of this discussion, the following anticipatory affective mechanisms were identified:

The representation of predicted states may be complemented (or even substituted in some processes) by the representation of an emotion and/or the representation of predicted emotions;

Predicted emotions and predicted representations must be confronted with sensed representations and emotions in order to produce the "current" affective state;
Predicted and current emotions as well as predicted and current representations should be used to predict a future affective state (Emotional Artificial Intelligence) as well as a future representation. Conversely, these predictions may affect the "current" emotions and representations;
The resulting anticipatory emotion(s) should elicit a preparatory behavior (the "tendency to act") that is "felt" by the system.
It is important that a framework be able of representing the result of such research. As such, the idea of developing a blueprint for an anticipatory affective system was laid, that also means a functional / procedural definition of anticipatory emotional mechanisms, build upon the enumerated elements. Additionally, the concept of parallel anticipatory paths (that all together form the "anticipatory continuum") was introduced, anticipating the need of a common ground for the affective integration in more general anticipatory systems. The next figure represents the "anticipatory continuum" idea.



ANTICIPATORY CONTINUUM

At a certain time, we have an expectation of a representation as well as the expectation of an emotion (both computed in the past). By confronting the predicted representation and the predicted emotion with the actual sensed/imagined representation and emotion, an affective state is generated. This affective state is then used to compute the expectation of an emotion and a representation of the future. The blue arrow in the figure represents the relation between emotion and representation, and will be implemented differently in each one of the architectures. Basically, we are adding a parallel path the to representation path of "plain" anticipatory systems, and within the Mind Races projects, the added value of this path will be evaluated.
Although the framework and its terminology have to be consolidated with empirical proof of concept, it is a first step towards an integration strategy to come and supports the mote that emotions are the glue of the past and future continuum, and do not exist without motion. The emotion is then the snapshot of this motion.

Description of the scenarios

We present here four scenarios that were discussed under the light of affective anticipatory systems. The idea is to transmit the rational that allowed us to reach the above conclusions, and present it within the context of the debate among the partners.

SCENARIO I

A robot is moving around in an environment with dangers that when met cannot be fully avoided (say fire). This danger leaves signs in the zone around its location (say: smoke or smell) that when encountered by the robot elicits some internal motion (or appraisal) in its body (i.e. a feeling of 'fright'). The robot learns to anticipatorily detect the danger just by conditioning an avoidance behaviour not to the danger but to its precursor sign.

In this scenario, there is no need of an explicit 'mental' representation of the future dangerous event. It is just a case where emotions elicit an anticipatory or preparatory behaviour to a possible event.
This seems to be for example the case of primitive forms of 'fright' where not a prediction or a belief about the future but simply the stimulus itself (like a noise or a sudden movement) elicits the emotion. We can ascribe this kind of emotion-based anticipatory behaviour to animals like rats or perhaps birds, but not to insects that only have merely reactive response. An example is particularly remarkable when considering Human beings: if a pregnant mother slipped and became frightened, the physical body state of the mother (pounding heart, muscle tension, etc) is experienced by the baby in her womb. The physical body states (associated with 'fright') are then stored in the baby's memory along with the perceptual context of the falling motion when its mother slipped. The same pounding heart will be experienced when, as an adult, the same person is flying in a jet airliner that experiences momentary turbulence. Same perceptual context, same physical body state, only this time it is experienced as a 'fear of flying'.
The problem of this rather simple scenario is that it is not at all clear what really the 'emotion' is in this case, what its function is, and why it shouldn't be simply skipped. Is the emotional 'mediation' just a trick, an empty and superfluous postulation? What does this add to the mere adaptive reactive behaviour?
First of all, the emotional response has a qualitative dimension, its experience is *pleasant* or *unpleasant*, and the feeling of this dimension provides the organism with an implicit 'evaluation' of the stimulus as *good* or *bad* for the organism. Such appraisal provides a sort of categorization of that kind for future uses that allows analogies and generalizations. For example, next time along the same path the robot might remember the felt fear and avoid this area without any sign of danger (smoke), just on the basis of the associated and evoked emotional experience.
Second, the emotional response plays a role of reinforcement in learning processes. It seems that this provides (independently of the success) an internal measure of the importance of the rule reinforcing it: the stronger the emotional activation the more reinforced the rule and the greater the probability that it will be activated next time in similar circumstances.
However, the lack of explicit 'mental' representation poses a problem: how does the

relation between emotion and preparatory behavior takes place? Supposing that it is 'inborn', which specific choices have to be made, namely which mechanism(s) are related to which emotion(s)? It was proposed that, as a general mechanism, the resource management strategies should take into account the current emotion when allocating processing resources. Note that the emotion processing itself needs a very low level of resources when compared to other cognitive processes.

Another problem that potentially makes the situation even more complex is related to a situation when a single stimulus causes several emotions. In such a case some selection mechanism must determine the possible reactions which could be different depending on the predominant emotional state. This is an issue that should be addressed and responded at the implementation level.

Furthermore, emotions should be remembered in context, that is, as related to places or environments. In this case, the situation (e.g. smoke, loud noise etc) elicits an emotion, which is related not to the reaction or the event but to the specific environment or context. In the future, the robot when in a similar context can remember its emotion and react in function of both the stimuli as well as the context.

Finally, to model something less empty, and closer to a bodily 'motion' and 'felt' emotion, a reactive variation of some internal bodily state will feed the learning mechanism.

SCENARIO II

The robot foreseen a given scene; this event is bad for it, is a threat, a danger. It feels 'fear' and changes its path (avoids the danger) or escapes away if the danger is moving and arriving. Later, while perceiving a possible danger or an unsafe zone or situation the robot, feeling a sense of anxiety, might multiply its investigating attitude and be more cautious but active for knowing about actual dangers or successes.

This scenario focuses on emotional responses that are caused by anticipatory representations, by predictions. In particular the robot feels its bodily reactions to endogenous representations of future events. Bad events, threats, elicit unpleasant emotions, while expected positive events elicit pleasant ones.

The function of this faculty should be an associative, intuitive, fast, experience-based appraisal of future events and situations impacting on current motivation, maintaining commitment against difficulties and procrastination, discouraging, or also activating other motives like reducing ignorance, acquiring additional information, being prepared.

A different relationship of the emotional response to the anticipatory representation is when the robot is facing the confirmation of disconfirmation of its expectations. The fact that there was a given prediction (mental representation of the future) where the organism was interested and concerned (that was important for its goals) and the fact that this expectation is invalidated or realized, elicit specific 'affective' states.

This is the area for the theory of 'Surprise'. The function of surprise seems to lie in the mobilization of resources for coping with 'abnormal' events (arousal), in particular processing/cognitive resources: attention. It seems that it is important for learning and changing habitual assumptions and rules: after a 'surprise' one cannot continue in its routine behaviour, must be aroused or careful.

SCENARIO III

The robot avoids a path (although perhaps this path is the shortest), because it predicts

– on the basis of the retrieval of previous experience and associated memory of felt emotions – that it will feel fear, and does not want to feel fear again. So the robot prefers certain paths because it expects to feel pleasure and joy there, although others may be shorter. Additionally the robot may anticipate another robot's emotions, and decide on its acting according to this affective prediction.

In this scenario the emotion is the object of the anticipatory representation, not its effect (I predict to feel guilt, or regret, or joy, or embarrassment) and it is taken into account in current reasoning or decisions.
The claim is that the ability to anticipate a possible emotion affects current decisions in various ways, and in particular (since emotions are positive or negative, i.e. one searches for them or wants avoid them) changes preferences about foreseen scenarios.
This scenario lays the foundation of Emotional Artificial Intelligence, as the robot can anticipate the emotions that a certain behavior can provoke, and act accordingly.

SCENARIO IV

When designing an agent system, we generally provide it with the ability to search the space into which it is integrated and devise an optimum plan allowing it to reach its goals while minimizing a cost function. Under this point of view, achieving a goal is one if not the most important concern of the agent and, of course, anticipation and anticipatory affect do play an important part in such design, as the other proposed scenarios will support. The main concern of scenario IV, however, follows a complementary approach: that the journey towards achieving the agent goal is as important as achieving the goal itself. This "zen" approach is especially relevant when designing believable synthetic character systems.
Consider the following example: Lucia throws a red ball into the next room, then turns to Aibo, the dog, and says: "Fetch!". Aibo runs into the room and designs a plan to find the red ball. While searching the space, its attention is drawn to a small handkerchief which color is just as the ball it is searching for. With its ear pointing forward, Aibo starts running, waving its tail and barking in anticipation. However, as soon as Aibo realizes it is a mere handkerchief, its ears drop back and its tail falls between its legs. With a disappointed face, Aibo starts moving back, its gaze wandering across the room...
From the planning algorithm point of view, Aibo may have found itself in a local minimum, however, from the user point of view, much more had happened, orthogonally to the search plan. When designing a system in which believability is a key factor defining the qualia of the interaction, the path can become more important than the goal itself. This is particularly important in terms of the end-user of anticipatory affective systems.

Scenario IV takes place in a household environment where Aibo, the synthetic dog, "lives". As a starting scenario, the environment will be a small warehouse, where several crates lie scattered around, acting as obstacles between Aibo and its targets. Several distracters, will be added to difficult the task and provide with opportunities for Aibo to "play in character" and be evaluated in terms of believability.
The believability evaluation will support the correctness of the affective eliciting mechanism.

6. A High-Level approach to Emotion and Anticipation

In this approach, we will focus on the close relationship between emotions and a basic anticipatory representation: goals. Emotions monitors and signal goal pursuit, achievement and failure; they generate goals; and finally they may translate into goals (Miceli and Castelfranchi, 2002a). Goals in turn are tied to beliefs, many of which are about future states or events.
We will present a cognitive symbolic approach for an affective anticipatory system, first addressing the route from emotion to anticipation, then the reverse one, from anticipation to emotion.

From Emotion to Anticipation

Trying to provide a route from emotion to anticipation we will start by distinguishing two "families" of emotions: *preparatory* and *premonitory* emotions.

Preparatory Emotions

Not any anticipatory behaviour is based on explicit cognitive representations of future events, that is, on predictions. Many instances exist of "implicit" or merely *behavioural anticipation* or *preparation*. This occurs whenever some stimulus that is *precursory* to a forthcoming event is associated with a certain behaviour, which has been selected to react to the forthcoming event (*preparatory behaviour*).
For example, the jumping of Alice at a rustle is not only a simple reaction to the noise itself, but it is (functionally) "meant" to avoid possible predators:

Precursory stimulus	⇒	Preparatory behaviour	⇒	Event
(noise)		(jump)		(approaching predator)

Often, the relationship between the precursory stimulus and the behavioural anticipation is mediated by emotions. A classical example is offered by fear, whose implied bodily activation is preparatory for flight behaviour. That is, the precursory stimulus elicits an internal emotional response, and the latter activates the anticipatory behaviour, which is preparatory for the forthcoming event:

Precursory stimulus	⇒	Emotion	⇒	Preparatory behaviour	⇒	Event
(noise)		(fear)		(flight)		(approaching predator)

From the perspective of biological evolution (e.g. Tooby and Coosmides, 1990), emotions are in fact psychological mechanisms that evolved to solve adaptive problems – such as escaping dangers – and predators, finding food, sheer and protection, finding mates, being accepted and appreciated among one's conspecifics – and thus surviving and delivering one's genes to one's own offspring. In other words, emotions generate goals and behaviours our ancestors had to pursue in order to answer such recurrent ecological demands. And, of course, the instrumental relation between such emotion-generated goals and their functions was far from being explicitly represented in our forefathers' mind.

There may be a variant of the preceding process, where no observable behaviour is present. In such cases the precursory stimulus elicits an emotion which is itself preparatory for the upcoming event. For example the fear elicited by a noise may just activate a state of vigilance and alertness, without any overt behaviour, and this very state is the preparatory response to the upcoming event:

Precursory stimulus	⇒	Preparatory emotion	⇒	Event
(noise)		(fear and alertness)		(approaching predator)

The question is: why did living systems evolve from an S ⇒ R schema to the S ⇒ E ⇒ R schema, with emotional (internal) 'mediating' response? If the function of the E response is just eliciting the overt behaviour B, why not have the simpler direct S ⇒ R association.[3]

The internal response E is likely to also play other roles. Mediating the relationship between S's and R's might allow that:

various stimuli S elicit the same internal reaction that elicits R and also, more important:

E plays a role in learning; it is reinforcement since it can be pleasant or painful. Finally, E remains associated to S in memory and is automatically retrieved – in a very fast and automatic way – during the perception of S (marker), and – in such a way – it represents *an implicit evaluation* of S (based on past experience) (Miceli and Castelfranchi, 2000a).

Premonitory Emotions

Emotions are *signals* of underlying mental states that account for and justify them. In other words, emotions accomplish an informative function. They provide some insight into oneself and one's relationship with the environment (e.g. Lazarus, 1991; Schwarz, 1990). Indeed we often realize and evaluate what is going on in a given situation, not before but after we experience some emotion. I can feel anger or fear, and then realize that something has happened that make me angry (someone has harmed me) or afraid (something is threatening me). To be sure, my interpretation may be incorrect, due to the ambiguity or vagueness of the emotional arousal combined with the interpretive bias favoured by a given context (e.g. Schachter, 1964). In any case, what we want to emphasize here is that emotions call for some interpretation, they demand some mini-theory about the reasons why I experience them.

Often, the very fact of experiencing a certain emotion or being in a certain mood (even, independent of external stimuli) elicits some anticipatory belief about a future state or event. For instance, my experiencing anxiety makes me suppose some impending danger; conversely, my cheerful mood this morning can induce me to feel that today is going to be a nice day. These are cases in which emotional states induce cognitive expectations about the future:

Precursory stimulus	⇒	Emotion	⇒	Expectation

In common usage, *expectation* is an ambiguous word. Sometimes it coincides with

[3] As it was pointed before, when dealing with anticipation, it would be possible to argue that it may be a question of parsimony and simplicity in terms of the organism function.

hope (or fear), sometimes with forecast, and sometimes it implies both. A simple forecast or *prediction* can be defined as a belief that a certain future event *p* is (more or less) probable, and it involves no necessary personal concern or goal about *p*. By contrast, by *expectation* here we mean an internally represented wish or goal about a future event together with the belief that the (un)desired outcome is possible or (more or less) probable. In other words, an expectation is a prediction the subject is personally 'concerned' about.

From Anticipation from Emotion

Also on the route from anticipation to emotion we can find two classes of emotion which are quite distinct from each other: expectation-based and invalidation-based emotions.

Expectation-Based Emotions

The expectation of *p* (that is, the anticipation or prediction of a certain event *p* together with the goal that *p* or not-*p*) is likely to elicit some emotion. For instance, if I expect failure at the exam, I will experience sadness and helplessness, or apprehension and anxiety; conversely, if I expect success, I will feel hope, joy, proudness, and so on. Here we are in the domain of cognitive appraisal proper (e.g., Lazarus, 1991), with the sole restriction that the appraisal regards *future* events.

$$\text{Expectation} \quad \Rightarrow \quad \text{Emotion}$$

Thus, whereas in premonitory emotions the latter induce some expectation, here the causal relationship between expectation and emotion is reversed. It is the expectation of a certain event (positive or negative, depending on its congruency with one's own goals) that elicits an emotional response.
However, some interesting cases may occur in which an originally expectation-based emotion turns into a preparatory or even a premonitory one. Let us sketch how.

From expectation-based back to either preparatory/premonitory emotions

Consider the following possible process: a stimulus elicits an expectation which induces an emotion, and the latter in turn activates a preparatory behaviour. For instance, suppose the following scenario: a child hears in the night a series of noises (stimulus) and recognizes them as those made by his father coming back home, usually drunk; such a stimulus, or better its recognition and evaluation, elicits the negative expectation (grounded on the child's previous experience in similar circumstances) that his father will thrash him; the expectation induces the emotion of fear, which in turn activates a preparatory behaviour, such as curling up in bed waiting for the thrashing:

Stimulus	⇒	Expectation	⇒	Emotion	⇒	Preparatory behaviour
(noise made by drunk father)		("he will trash me")		(fear)		(curling in bed)

Through habituation this process may undergo a "short-circuit", where there is no longer any explicit expectation: that is, the emotion comes to be *directly* triggered by the stimulus:

Stimulus	⇒	Emotion	⇒	Preparatory behaviour
(noise made by drunk father)		(fear)		(curling in bed)

In this way, an expectation-based emotion can turn into a preparatory one. Moreover, the process may undergo some generalization: for instance, similar noises in similar contexts may directly trigger the same emotion (fear), independently of any explicit expectation. What was originally based on an explicit expectation and now has turned into a preparatory one.
Interestingly enough, such a preparatory emotion, if consciously felt and reasoned upon (why am I feeling this terror? Why am I feeling this urge to hide?) may favour, in some cases, the retrieval of the forgotten expectation. In other cases, it can generate a generic expectation of threat: "since I am scared, there must be some impending danger somewhere". In this way, the preparatory emotion turns into a premonitory one, where, as already pointed out, a stimulus elicits an emotion, the emotion induces an expectation and the latter, in turn, a preparatory behaviour:

Stimulus	⇒	Emotion	⇒	Expectation	⇒	Preparatory behaviour

Invalidation-based Emotions

Suppose that at time t_1 I have a certain (positive or negative) expectation, and that at time t_2 my expectation is falsified. The fact of having had an expectation and its being invalidated produces some emotion. If the expectation was *positive* – that is, my goal was congruent with my prediction – I experience disappointment. For instance, if I both predict and want that John comes and sees me (or I both predict and want that he does not come) and I find my expectation invalidated by actual facts, I will be disappointed. Conversely, if my expectation was a *negative* one – that is I wanted *p* and predicted not-*p*, or vice-versa – and it is invalidated, I experience relief.

Expectation (t1)	+	Perceived event (t2)	⇒	Expectation invalidation	⇒	Emotion

A number of relevant remarks are worth making in this connection. First of all, it should be stressed the crucial role played by the anticipatory belief and its invalidation in these contexts. In fact, such emotions as *disappointment or relief cannot be elicited without anticipatory beliefs*. Mere goal compromising or fulfilment, if devoid of any specific prediction (e.g., I want John to come and see me, but I do not make any particular forecast on this matter), can of course elicit some emotion (either pleasant, such as joy, or unpleasant, such as sadness). But disappointment proper can arise only if my goal was accompanied by a (more or less certain) *prediction* about its fulfilment, and this prediction has been invalidated. In the same vein, I cannot feel relief unless I *predicted* some goal thwarting that does not come true.

Predictions on the Emotion's Intensity

Given the cognitive ingredients we postulate in these 'invalidation-based' emotions, we assume that the intensity of the emotion is function of its components. In particular:

(i) *The more (subjectively) certain the prediction, the more intense the disappointment or the relief.*

(ii) *The more (subjectively) important (valuable) the goal, the more intense the disappointment or the relief.*

Expectation validation and emotions

We have pointed the relationship between expectation invalidation and emotion but what can we say about the *validation* of expectation? Does any specific emotion depend on the validation of one's own expectations? We do not suppose any remarkable qualitative difference in emotion elicitation between a case in which a mere goal (without prediction) is fulfilled or thwarted, and a case in which an expectation (either positive or negative) is validated. To be more precise, we assume that the possible difference lies in the *intensity* of the emotions experienced rather than in their *quality*. As a general rule, we suggest that, *if compared with the emotions elicited by mere goal (without prediction) fulfilment or thwarting, those emotions which are elicited by validated (either positive or negative) expectations should be lower in intensity*. That is, the pre-existing prediction plays the role of "watering down" the (positive or negative) emotion associated with the destiny of the goal. In fact, expected outcomes have lower emotional impact than unexpected ones: As expected negative outcomes are less painful than unexpected ones, so expected positive outcomes are less elating than surprising ones (e.g., Mellers, Schwartz, Ho and Ritov, 1997; Miceli and Castelfranchi, 2002b).

Kinds of invalidation-based emotions

Going back to expectation invalidation, its impact on the emotional system is not limited to such feelings as disappointment and relief. On the negative side, at least a couple of other feelings are worth mentioning: discouragement and sense of injustice.

Disappointment vs. discouragement

Disappointment implies a process of transition or transformation of a positive expectation into a negative one. A disappointed expectation is in fact a positive expectation (with varying degrees of certainty) that becomes negative (with varying degrees of certainty). Discouragement is a particular kind of disappointment whereby one's disappointed expectations concern one's own (either internal or external) power to realize an intention (Miceli & Castelfranchi, 2000b). Also in discouragement there is a transformation of a positive expectation into a negative one. In particular, discouragement implies a transition from a situation where one has the "courage", that is, one feels to "manage" it, to a situation where one loses heart, and feels not to "manage" it, that is, one comes to despair of achieving some goal after having expected a positive outcome.

However, a discouragement is something more specific than a simple disappointment. Discouragement implies disappointment, whereas there may be disappointment without discouragement. Suppose yesterday I expected to have a sunny weather today: if today my expectation is invalidated, I get disappointed, but not discouraged. In this case, there is nothing to be discouraged about. In fact, one may be disappointed about mere goals, while, for being discouraged, there should necessarily be some intention (that is, some goal chosen for pursuit) implied: one is discouraged from pursuing some goal (because one's positive expectations have been disappointed). Going back to the previous example, discouragement might come into play if the expected sunny weather were considered an enabling condition for pursuing the goal of, say, taking a trip. In such a case, I would be discouraged with regard to that goal (or better intention), while I am just disappointed relatively to the goal of having a sunny weather.

Moreover, discouragement shows another important difference from mere disappointment. In discouragement the focus of attention is put on one's *lack of (either internal or external) power* to achieve a certain intention *p*, whereas disappointment is, so to say, *unmarked with regard to power*.
Though both imply a transition from a positive expectation to a negative one, in the case of discouragement the positive expectation consisted in a belief of the type "I can manage it", while in the case of mere disappointment it could just be "*p* will happen". This is quite in line with Weiner, Russell and Lerman's (1979, p. 1216) view of disappointment as "independent of attributions but dependent on outcomes" (see also Zeelenberg et al., 2000).

Sense of injustice

Sense of injustice is a likely response to invalidated positive expectations. The stronger the positive expectation (that is, the more certain its implied prediction are and the more important its implied goal is) the more its invalidation subjectively looks like an *ill-treatment*, as if one were suffering something *unfair*. In fact, anger is a common reaction to a violated positive expectation (Averill, 1982; Burgoon, 1993; Levitt, 1996; Shaver, Schwartz, Kirson, and O'Connor, 1987), as well as to perceived unfairness (Fehr and Baldwin, 1996; Fitness and Fletcher, 1993; Shaver, Schwartz, Kirson, and O'Connor, 1987). The assumed violation is accompanied by a sense of rebellion and refusal of facts (in fact, they 'shouldn't have gone' as they did). What I expected resembles what I was *entitled* to obtain. I feel I didn't *deserve* what has happened (Miceli & Castelfranchi, 2002b). This feeling of injustice is somewhat metaphorical in that no *explicit* subjective equation of 'expected' with 'deserved' is necessarily implied. There is just a sort of implicit and analogical overlap of the two concepts. The reason for this implicit overlap lies in a special normative component typical of positive expectations which is absent from the other kinds of anticipatory representations of the future. Positive expectations in fact do not simply consist in 'predictions plus goals': they also imply a normative component, which results from the translation of the epistemic normativity typical of predictions into a deontic normativity: What, in probabilistic terms, 'should' happen, and I want to happen, turns into what I am entitled to obtain.

But why should the epistemic 'norm' be made equal to the deontic one? Because *a positive expectation favours an 'as if' state of mind, according to which the desired state is viewed as (almost) realized*, and the individual feels already allowed to enjoy its satisfaction. Therefore the realization of the goal is represented as something to be *maintained* rather than acquired. Because a maintenance goal (as opposed to an acquisition one) is likely to be viewed as grounded on some supposed right (a sort of usucapion), people feel entitled to obtain what they expect (Miceli & Castelfranchi, 2002b). In other words, the relationship between maintenance goals and positive expectations can account for the ease of translation of the epistemic norm into the deontic one.

Another, more general, reason for such a translation lies in the common tendency to turn mere implications into equivalences. Because perceived rights create positive expectations, we are also likely to surreptitiously assume that positive expectations create some right! As often happens in everyday reasoning, conditionals 'invite' the biconditional interpretation (e.g., Geiss & Zwicky, 1971; Oaksford and Stenning, 1992; Wason and Johnson-Laird, 1972), and simple implications ('if *p* then *q*', that is, 'if there is a right, there is a positive expectation') are turned into reciprocal ones ('if *p* then *q*' *and* 'if *q* then *p*'), i.e., equivalences. As a consequence, 'if there is a positive expectation, there is a right'.

The case of invalidated negative expectations: surprise and relief

The sense of injustice that is typical of disconfirmed positive expectations does not seem to be experienced when negative expectations are disconfirmed. The reason for this difference lies in our view in the absence of a normative component in negative expectations. Actually, when I want *p* but I predict not-*p* (or vice versa), I do not set any deontic norm that *p* or not-*p* ought to happen. I just believe, on the grounds of my experience or previous knowledge, that not-*p* is likely to happen, whereas I would prefer the opposite. When my negative expectation is disappointed, of course I will be *surprised*. But my surprise will take on a positive colour, because my goal *p* has been fulfilled. I will neither protest nor look for somebody's responsibility, nor feel I have been treated unfairly. Rather, I will feel *relieved*, because, contrary to my prediction, my desire is fulfilled. Actually, before an unexpected happy ending the typical feelings are surprise and relief. The latter will be all the greater the more important is the goal fulfilled, and the more unexpected its fulfilment. Relief in fact implies a more or less explicit comparison between the anticipated distress and the actual positive situation.

Thus, a *normative component is implied only in positive expectations*. This amounts to saying that the normative component results from the *joint* force of predictions and goals. If predictions and goals are congruent with each other, then *p* 'ought' to occur. If they do not converge (I predict not-*p* and I want *p*, or vice versa) no normative component will be implied. A negative expectation, when invalidated, is just *disappointed*, whereas a positive expectation, when invalidated, is *violated*.

Prediction invalidation and emotion

So far, we have considered the emotional responses associated with the invalidation of *expectations* proper, that is, predictions *plus* goals. However, not only expectation proper, but also mere *prediction* invalidation (that is a disconfirmed forecast that *p* devoid of any goal that *p* or not-*p*) may elicit some emotion. This is, again, the case of *surprise*.

As just remarked, mere predictions do not imply any personal concern about *p*, in the sense that *p*'s occurrence does not affect any of the person's goals. For instance, my prediction that next Wednesday John will visit Mary (because this happens each Wednesday) may have nothing to do with my goals: I have no interest in the fact by itself. In this sense, a prediction is a 'cold', or better neutral, belief that 'probable *p*'. However, if this 'neutral' *p* does not occur, we are likely to experience a surprise which contains a certain degree of distress, as if we *wanted* *p* to become true. The more certain the prediction, that is, the more *p*'s assumed probability is close to 100%, the more the surprise turns into a *bewilderment* that is tinged with a negative connotation.

But, if we do not have the goal that *p* by itself, what is the 'goal' implied in a prediction?

People have a need for prediction, that is, they need to know what causes will come into play to produce what effects (whether beneficial or harmful). The need for prediction implies both a need to anticipate future events and the consequent need to find such anticipations validated by facts. This is Bandura's (1982) *predictability*, i.e., the cognitive component of *self-efficacy* (as distinct from the other component, *controllability*, i.e., the need to exert power over events). However we assume that the need for prediction is *not* a goal proper, that is, it is *not* a regulatory state represented (consciously or unconsciously) in the person's mind, but a *meta-goal*, that is, a *regulatory principle* concerning one's mental functioning (Miceli and Castelfranchi, 1997). Consider belief consistency. In a sense, we 'want' to maintain consistent beliefs. In fact, if a contradiction is detected, we try to eliminate it. However, the mind has this 'goal' as a function. It is not necessary to express these finalistic effects as represented goals on the basis of which the mind reasons and plans. It is sufficient to conceive these principles as *procedures*, which are implemented when a contradiction is detected. If they are unsuccessful, a form of cognitive distress is likely to be experienced. In the same vein, the mind's architecture includes the meta-goal to make predictions and to find those predictions validated by the evidence. Finally, the meta-goal to find one's predictions validated implies the further meta-goal that *p* happens (since, according to one's beliefs, it should happen). This can account for the surprise experienced and its likely negative connotation, which is stronger the more certain the prediction, and comes close to a sense of bewilderment, because the world is less predictable than expected. This view can also account for the tendency to behave in accordance with one's predictions in those cases when one's behaviour can affect the likelihood of the predicted event (see Sherman, 1980).

Expected Emotions

A third general case of interaction between expectations and emotions is offered by explicit representations of *future states which coincide with emotions*. In other words, emotions are here the *object* of anticipatory representations, rather a reaction to them: "if I do *a*, I will feel guilty" (or happy, ashamed, relieved; and so on). Two kinds of expected emotions can be identified: 'cold' expectations and 'hot' expectations of emotions or, better, *expected and non pre-felt emotions* versus *expected and pre-felt emotions*. The latter include some *anticipated feeling*. In both cases, expected emotions play a remarkable role in the decision-making process: expecting possible emotions as a consequence of one's candidate decisions affects the latter, changing one's preferences about the given options.

Expected and non pre-felt emotions

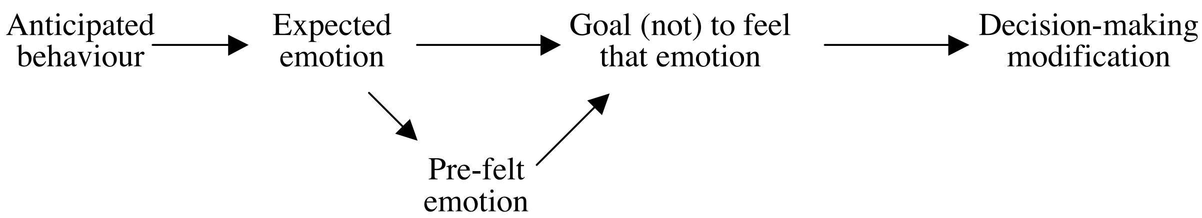
By 'expected and non pre-felt emotions' we mean those emotions the agent predicts to feel as a consequence of a candidate decision, but that he is not actually feeling 'here and now'. The main point to be remarked is that a 'not-yet-felt' but *expected* emotion can enter the overall evaluation of what goals are worth pursuing, adding a new way of linking emotions to decision making. It is worth noticing that though expecting to feel an emotion is sufficient for changing the decision process or its results, the agent does not have to feel it either now or later (that is, his prediction may be wrong).

Expected emotions belong to the set of tools an agent can use for discriminating among different choices, so that he or she can evaluate which choice leads to the best outcome, including the emotions he or she would like to feel, or at least those he or she would be more able to stand (in the case of choices implying unpleasant emotions). In other words, while anticipating some future course of action, the agent is also likely to anticipate that he or she would feel some particular emotion; this (positive or negative) expected emotion induces the goal (not) to feel it, and this goal enters the decision-making process with a given value, possibly modifying the value of the available options. For instance, while considering how to obtain advancement at work and anticipating some way for cheating a colleague of mine, I expect to feel guilty; and this expectation can induce the goal not to feel guilty, to such a point that I give up the option of cheating:

Anticipated behaviour ⇒ Expected emotion ⇒ Goal (not) to feel that emotion ⇒ Decision-making modification

Expected and pre-felt emotions

The expected emotion can also be pre-felt: While anticipating a possible behaviour and its context, I am in fact likely to 'foretaste' the emotion I expect to feel, at least to some degree of intensity (if not to the same degree as when the anticipated situation is actualized). For instance, going back to the previous example, I may feel guilty at the prospect of cheating my colleague, that is, I may 'hallucinatorily' experience what I (believe I) would feel if I cheat my colleague. In such cases, the impact of the expected emotion on decision making is probably stronger than in expected but non pre-felt emotions. In fact, here the mere cognitive expectation about some emotional reaction is reinforced by its 'foretaste':

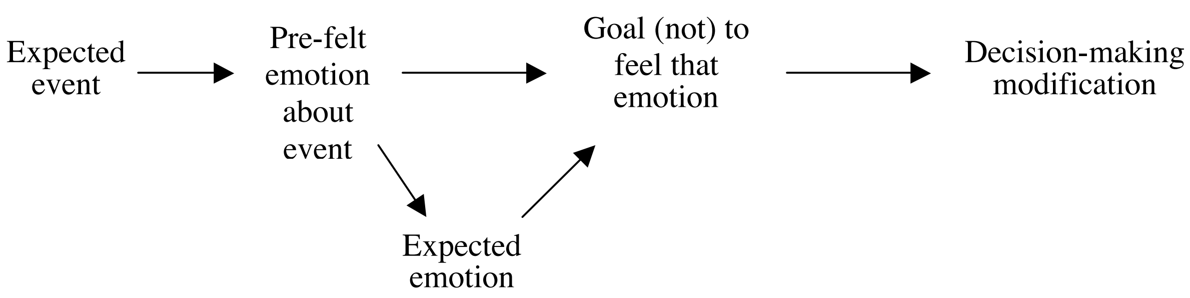


Sometimes, the expectation that one will feel a certain emotion may elicit an emotion which is different from the expected one. For instance, at time t_1 I may feel *fear* at the prospect of feeling guilty at time t_2. In such cases, the emotion experienced at time t_1, rather than being a foretaste of the expected emotion, is an expectation-based emotion (see above) in the strictest sense. In comparison with the expectation-based emotions we have already considered, here the difference lies in a further specification: the expectation concerns an emotion (the emotion I will/would feel at time t_2). Thus, an expected emotion may both favour the foretaste ('pre-feeling') of the emotion expected, and elicit some other emotion *about* the expected emotion. And in any case, such feelings are likely to impact on the decision-making process.

A hybrid case, which we might call 'expectation-elicited emotion' is the following: An expected event ("the boss will fire me") elicits an emotion about it (anger), and this emotion 'tells' me what I will probably feel when the event happens, that is, the emotion is the evidence on which I ground my expected emotion. This is an interesting case, for at least a couple of reasons:

On one hand, it resembles the process implied in the expectation-based emotions, in that also here an emotion is elicited by an expectation; however, whereas in the expectation-based process the emotion is experienced as regarding a *future* event (e.g., I *now* feel hope, fear, disappointment, discouragement at the *prospect* of a certain outcome), an 'expectation-elicited emotion' implies a sort of *actualization* of the future event: while imagining my boss firing me, I feel (an amount of) the anger I will/would feel at that point in the future, when/if that event happens;

On the other hand, expectation-elicited emotions are akin to expected and pre-felt emotions, in that in both cases pre-feeling the expected emotion impacts on the agent's decision-making. But, whereas in the expected pre-felt emotions an expectation *about* an emotion favours its foretasting, in the expectation-elicited emotions, it is the other way around: foretasting favours the expectation that I will feel that emotion (since I am pre-feeling it now); that is, foretasting plays a *premonitory* role:



Concluding remarks

We have analyzed and systematized:

- The *role of emotions in/for anticipation* both in terms of eliciting/confirming (non-neutral) anticipatory mental representations (*expectations*, or in eliciting an anticipatory (preparatory) behaviour;
- The role of anticipation in different kinds of emotions:
- Some of them are feelings joined to expectations (like hope or fear);
- Some are the result of the frustration or confirmation of expectations (like surprise, disappointment, relief);
- Some are the anticipated representation (and possibly also feeling) of a future emotion.

To conclude let us just stress that the relationship between Emotion and Anticipation is neither of overlap nor of inclusion. They are just *partially* overlapping sets of phenomena.

On the one side, the capacity for anticipation is not necessarily emotion-based. Non-emotional systems can be 'anticipatory', not only behaviourally but also cognitively; that is building internal representation of future events (predictions and expectations). On the other side, Emotions are not necessarily based on anticipatory representations, nor necessarily anticipated.

However, the capacity for anticipatory representations:

- Creates new emotions (from hope and anxiety to disappointment);
- Creates the possibility for anticipated/predicted emotions (ex. fear of possible blushing);
- Emotions can produce both anticipatory behaviours and anticipatory mental representations.

6. A Low-Level approach to Emotion and Anticipation

An agent system generally focuses all its resources towards achieving a set of goals. The approach presented below follows a complementary approach: that the journey towards the goal is as important as the goal itself. This approach is particularly relevant when designing believable synthetic character systems. This section presents a simple extension to the base agent architecture: an anticipatory module that (1) monitors the information flowing back and forth between the different agent modules; (2) tries to anticipates which information is likely to be monitored in the future and; (3) by confronting the real information with the anticipated prediction, using a model inspired in Emotion and Attention research, provides with a simple automated sensation and attention control mechanism that, besides providing with guidelines towards human-like resource management, also provides the agent system with an autonomous mechanism to enhance its believability while acting towards its goals.

Zen of Anticipation

When designing an agent system, we generally provide it with the ability to search the space into which it is integrated and devise an optimum plan to reach its goals while minimizing a cost function. Under this perspective, achieving a goal is one if not the most important concern of the agent. This section proposes a complementary approach, arguing that the journey towards achieving the agent goal may be as important as achieving the goal itself. This approach is particularly relevant when designing synthetic character systems.

Consider the following example. Mickey throws a red ball into the next room, then turns to Pluto and says: "Fetch!". Pluto runs into the room and designs a plan to find the ball. While searching the space, its attention is drawn to a small piece of plastic that looks just like the ball it is searching for. With its ears pointing towards the red stimulus, Pluto starts running, waving its tail and barking in anticipation. However, as soon as Pluto realizes it is a mere piece of plastic, its ears drop back and its tail falls between its legs. With a disappointed face, Pluto slowly starts moving back, its gaze roaming across the room. From the planning algorithm point of view, Pluto may just have found itself into a local minimum, however, from the user point of view, much more had happened, orthogonally to the search plan. When designing a system in which believability is a key factor defining the qualia of the interaction, the path can become more important than the goal itself.

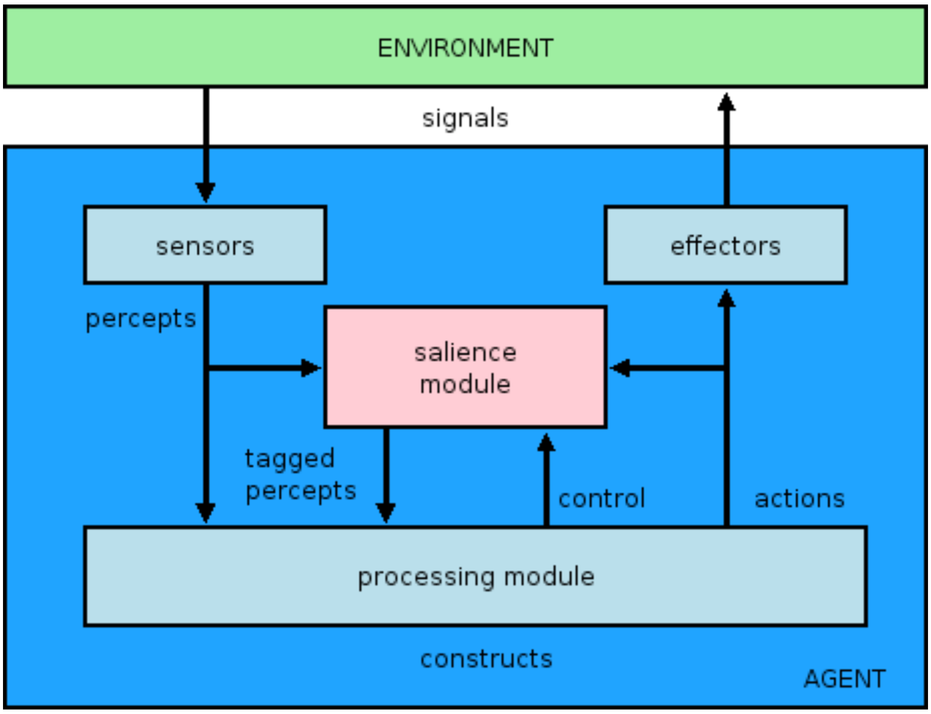
The next sections present an extension to the base agent architecture that follows this "zen" approach. They first introduce the concept of believability and describe the behaviour loop, a simple affect-based control strategy used to enhance synthetic characters believability. Then will follow a description of how the base agent architecture can be extended with an autonomous module with anticipatory capabilities, that monitors the information flowing between the different agent modules and generates affective data that can be used to implement the behaviour loop. Afterwards, the text delves into the concept of emotivector (the base component of the anticipatory module), explains how emotions and attention are integrated in its control, and describes strategies to manage several emotivectors at once.

Autonomous Believability

Synthetic characters are an effective medium to enrich the interaction between the user and the machine. Upon defining the qualia of the interaction with synthetic characters, a critical yet subjective factor appears: **believability**. By a believable character, we mean a digital being that "acts in character, and allows the suspension of disbelief of the viewer" (Bates, 1994).
Disney animators have been designing believable characters for almost a century and have developed a set of guidelines to help in the creation process (Thomas and O. Johnson, 1994). Most guidelines are based upon the principle of "showing the internal state to the viewer". By making clear what the relations between the synthetic character and its environment are, the character is made **aware** of its surroundings: even the synthetic characters are not "real", the relations between them are!
The concept of awareness can be further developed into what we call the **behaviour loop**: agents should change the focus of their attention and respond emotionally to stimuli provoked by other agents, and these reactions should be responded to as well. As an example, when Mickey enters the room, Pluto should react by looking at him and clearly expressing an emotion, perceived as being caused by Mickey. In response, Mickey should look and express an emotion back to Pluto. This loop is a simple strategy that increases the believability of the intervening characters.
The approach presented here propose a set of mechanisms that are suited to control both the focus of attention and the emotional reactions of a synthetic character in real time, to increase its believability through the behaviour loop and, trying to provide this control in a form as independent from the main agent processing as possible. All in the attempt of creating mechanisms for "autonomous believability".

Architecture

Our synthetic characters are implemented as software agents (Russel and Norvig, 1995). To make the control as independent as possible from the agent processing, we provide the agent with an autonomous module: the **salience module** (see next Figure).



The salience module performs a semantic independent monitoring of the percepts flowing from the sensors to the processing module as well as the action-commands flowing from the processing module to the agent effectors. This monitoring is possible since the code of the information flowing through the agent is usually consistent, in the sense that it is the repeated measurement of a specific internal or external aspect of the agent on a same scale over time. As such, we assume our universe of perceptions to be an n-dimensional vector space where each aspect of the world is normalized to the range [0, 1].

The salience module is composed of several emotivectors, each one associated with a sensed dimension. An emotivector is a module that keeps a limited record of a signal history and possesses mechanisms to estimate the next expected value based on this history. By confronting the expectation with the sensed value, and using the affective model explained in next section, the emotivector computes the sensor salience, and the percept is tagged with information providing both its attention focus potential as well as its emotional potential. The salience module also possesses a set of strategies to manage all the emotivectors together. The strategies will be discussed later.

The tagged percepts reaching the processing module of the agent are meant to be used as a guideline towards human-like behaviour. For instance, they can be used as parameters for an autonomous mechanism controlling the synthetic character behaviour loop.

Affective Model

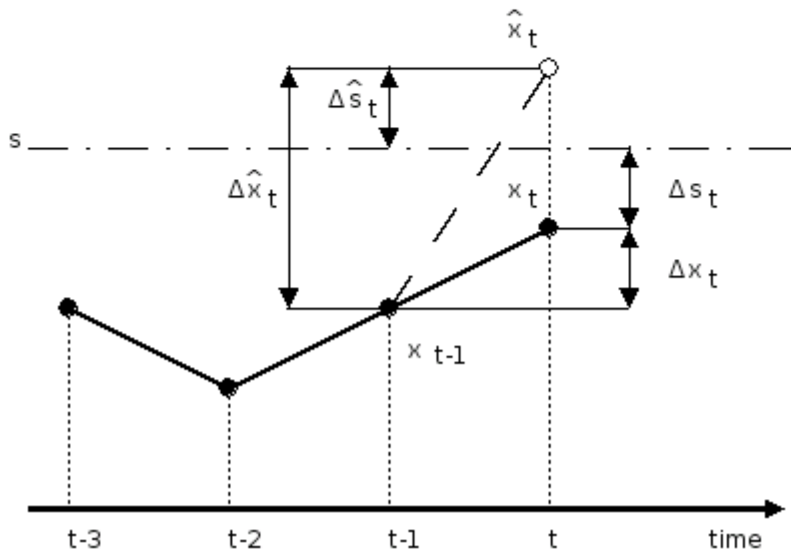
Rather than trying to implement a detailed affective model (Picard, 1997), this approach selected a small set of principles from the Psychology of Attention and Emotions. Even these principles fail in providing with an accurate description of how we Humans act, they are useful in building simple synthetic models of behaviour that perform well in real time, a critical issue when considering the creation of "autonomous believability". As emotions and attention cannot be considered separately (Wells and Matthews, 1994), these simple principles were merged into one combined approach. The next subsections describe our attention and emotion model, first individually and then, integrated as one combined affective model.

Attention

Posner (Posner, 1980) showed that directing attention to a valid location facilitates processing, and this led him to suggest that "attention can be likened to a spotlight that enhances the efficiency of the detection of events within its beam." Thus, it seems, attention is oriented to a stimulus. Posner experimented with central and peripheral cues and found that the attentional spotlight could be summoned by either cues, but peripheral cues could not be ignored whereas central cues could. Posner proposed two attentional systems: an **endogenous** system, controlled voluntarily by the subject and an **exogenous** system, outside of the subject control, which automatically shifts attention according to environmental stimuli and cannot be ignored.

While performing a series of experiments to clarify Posner's hypothesis of two distinct attentional orienting mechanisms (rather than only one mechanism controlled in different ways), Muller and Rabbit (Muller and Rabbit, 1989) showed that exogenous orienting could sometimes be modified by voluntary control. They suggested the following hypothesis, compatible with Posner's proposal: "Reflexive orienting is triggered and proceeds automatically, and if both reflexive and voluntary orienting mechanisms are pulling in the same direction, they have an additive effect. However, if they are pulling in different directions, their effects are subtractive."
Following Posner's theory, the model present here uses two interacting components to compute the emotivector salience: an exogenous component and an endogenous component.

At step *k-1*, the emotivector value is x_{k-1} in [0, 1]. Using its history at time step $k-1$, the emotivector estimates a value for next time step k (x'_k) and predicts that its value will change by $\Delta x'_k = x'_k - x_{k-1}$. At step k, a new value is sensed (x_k) and a variation $\Delta x_k = x_k - x_{k-1}$ is actually verified (see next Figure).



The exogenous component (EXO_k) is based on the estimation error and reflects the principle that the least expected is more likely to attract the attention:

$$EXO_k = (x_k - x'_k)^2$$

The endogenous component (END_k) is computed whenever a search value is given to the emotivector.

The exogenous component is mainly attentional. The emotional aspect of exogenous control is not considered at this stage but will be integrated as described in (Martinho, 2003). It is a function of the variation of the distance to the search value (Δs_k) and its estimated value ($\Delta s'_k$):

$\Delta s_k = (x_k - s_k)^2$
$\Delta s'k = (x'_k - s_k)^2$
$ENDk = \Delta s'_k - \Delta s_k$

Unlike EXO_k, which is always nonnegative, END_k is valenced: an increase in the expected search distance is assumed negative, while a reduction is modeled by a positive value.

Both added exogenous and endogenous salience define the relevance of the emotivector, following Muller hypothesis. However, an emotivector with a search value also possesses a certain qualia. This is described in the next section.

Emotion

As Harlow and Stagner (Harlow and Stagner, 1933), we differentiate between sensation and emotion. Harlow and Stagner proposed that there are basic sensations, innate and undifferentiated, and that emotions are a conditioned form of these sensations, which we learn to refer in particular ways. We are born with the capacity to feel but have to learn the different emotions. As Harlow and Stagner, we assume that emotions are conditioned responses of primary sensations, and concentrate our model in the generation of these sensations. Emotions per se are left to the processing module cognitive or symbolic affective processing.

As Young (Young, 1961), we do not speak of emotions but of affective processes in a hedonistic continuum. Hedonistic changes can occur in either positive or negative direction, giving form to four possible affective changes: positive increase, positive reduction, negative increase and, negative reduction. As Young, we give to the affective processes a motivational and regulatory role driving, among other things, the subject toward or away from a stimulus.

Inspired by the behavioral synthesis of Hammond (Hammond, 1970), we consider the sensation as a central state of the organism which is generated by stimuli, both known and unknown, that relates to the presence or absence of a reward or punishment. We use the emotivector estimation to anticipate a reward or punishment which, when confronted to the actual value, triggers one of Hammond's four basic sensations (fear, relief, hope and distress) that we translate to Young's sensations.

Inspired by Millenson (1967), we attribute an intensity to each sensation, which value is the emotivector salience, and allow a same sensation to vary across a certain range. We also use Millenson designations for our sensations, as the symbols are not connoted to an exact word which, by itself, would imply a certain intensity.

To sum up, we consider the following five sensations:

surprise (S) when there is no expectation of a reward or punishment due to the absence of a search value in the emotivector.

positive increase (S+) that we relate to Harlow and Stagner's excitement as well as to Hammond's hope and Millenson positive unconditioned stimulus, and associate with a reward stronger than the expectation. If reward is anticipated and the effective reward is stronger than the expected, a S+ sensation is thrown.

positive reduction ($+) that we relate to Harlow and Stagner's discontentment as well as to Hammond's distress and Millenson reduction of a positive unconditioned stimulus provoking rage, and associate with a reward weaker than expected. If reward is anticipated but the effective reward is weaker than the expected, a $+ sensation is thrown.

negative increase (S-) that we relate to Harlow and Stagner's depression as well as to Hammond's fear and Millenson negative unconditioned stimulus provoking anxiety, and associate with a punishment stronger than expected. If punishment is anticipated and the effective punishment is stronger than expected, a S- sensation is thrown.

negative reduction ($-) that we relate to Harlow and Stagner's pleasure as well as to Hammond's relief and Millenson reduction of a negative stimulus, and associate with a punishment weaker than expected. If punishment is anticipated but the effective punishment is weaker than expected, a $- sensation is thrown.

The expected reward at time step *k* (R'_k) and the sensed reward (R_k) are computed as follows:

$R'_k = \Delta s_{k-1} - \Delta s'_k$

$R_k = \Delta s_{k-1} - \Delta s_k$

Anticipation

The computation of the emotivector salience rely on the capacity of the emotivector to predict its next state. Before anything else, we define the model that we expect the sensed data to follow – if the signals are totally random, no prediction strategy can be evaluated for adequacy.

Model

As there is no a-priori knowledge of the signal, we follow a simple assumption: that the intensity *i* of a signal will change by a random small amount Δi in [−e, e] at each discrete time step (defined by the sensor rate), for a random time slice Δt in]0, Δt_{max}], before suddenly changing to a random new value in the interval [0, 1]. In other words, the sensed value will tend to remain constant except for certain points in time. By modulating the two model parameters e and Δt_{max} across an acceptable range and measuring the accumulated squared error over a series of samples, we verified the adequacy of the predictors described in the next sections.

Polynomial Extrapolation

A first consideration would be to use polynomial extrapolation. The idea was to find the polynomial P of degree N−1 interpolating the N points of the emotivector history and then use this polynomial to compute the next state. As low-degree polynomials give too little flexibility in controlling the shape of the curve and higher-degree polynomials can produce unwanted wiggles and also require more computation, we settled for cubic polynomials associated with a buffer memory of 4, inspired by the parametric curve approach (Foley, van Dam, Feiner and Hughes, 1990). After experimenting with several curves, we found that Uniform Non-Rational B-Splines and Catmull-Rom Splines were a possible choice, as they behaved well and had an acceptable performance. However, the adaptation to the sudden changes in the model is very slow, which led us to look for other solutions.

Error Driven Learning

The second approach was based on the Least Mean Square (Widrow and Hoff, 1960), also known as the delta rule. The LMS makes direct use of the discrepancies or errors in task performance to adjust the mapping weights between input and output and minimize the error measure. The base equation is essentially the same equation used in the Rescorla-Wagner rule classical (Pavlovian) conditioning (Rescorla and Wagner, 1972). Although this learning as being labelled as supervised learning, there are a multitude of valid sources of such "supervision" that do not require the constant presence of an omniscient teacher. The multilayer generalization of the delta rule is called back-propagation, which allows error occurring in a distant layer to be propagated backwards to earlier layers. Although the original mathematically direct mechanism for implementing the back-propagation algorithm is biologically implausible (O'Reilly and Munukata, 2000), an algorithm called **recirculation** (Hinton and McClelland, 1988) provided with ideas that enabled back-propagation to be implemented in a more biologically plausible manner. The algorithm has two activation phases: the **minus** phase, where the outputs of the system represent the expectation of the system as a function of the standard activation settling process in response to a given input pattern and; the **plus** phase, where the environment is responsible for providing the target output activation. The learning is essentially just the delta rule confronting the expectation with the sensed value:

$\Delta w_{ik} = LR\ (y^{+}_{k} - y^{-}_{k}).s_{i}$

where Δw_{ik} is the weight adjustment for a receiving unit with activation y_k (y^{+}_{k} and y^{-}_{k} in the plus and minus phases, respectively) and a sending unit with activation s_i using a learning rate of *LR*.

Kalman Filtering

A Kalman filter (Kalman, 1960) is a set of mathematical equations that provides with an efficient computational recursive means to estimate the state of a process in a way that minimizes the mean of the squared error. The filter is composed by two sets of equations: the **time update** equations, that project forward (in time) the current state and error covariance estimates to obtain the a-priori estimate for the next time step and; the **measurement update** equations, that are responsible for the feedback i.e. for incorporating a new measurement into the a-priori estimate to obtain an improved a posteriori estimate.

Following our signal model, a Kalman filter estimating a random constant was implemented using the following equations, where *R* is the measurement noise covariance and *Q* is the process noise covariance.

Time update equations

Project the state ahead

$x'^{-}_{k} = x'_{k-1}$

Project the error covariance ahead

$P^{-}_{k} = P_{k-1} + Q$

Measurement update equations

Compute the Kalman gain

$K_k = P^{-}_{k} / (P^{-}_{k} + R)$

Update the estimate with measurement

$x'_{k} = x'^{-}_{k} + K_k\ (x_k + x'^{-}_{k})$ (Eq. 1)

Update the error covariance

$P_k = (1 - K_k)\ P^{-}_{k}$

Simple Predictor

There are obvious similarities between the concepts presented: the two phases of the recirculation algorithm follow the same principle as the two steps of the Kalman filter algorithm and both algorithms need a certain parameter tuning in order to work: the noise covariances Q and R in the case of the Kalman filter and the learning rate LR in the case of the delta rule.

Following the same two-phases principle, we developed a simple predictor free from parameters. We started by rewriting x'_k from Eq. 1 as:

$x'_k = x^-_{k-1} \cdot R / (P_{k-1} + Q + R) + x_k \cdot (P_{k-1} + Q) / (P_{k-1} + Q + R)$

And then made the following substitutions,

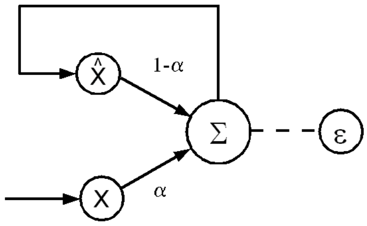
$\alpha = (P_{k-1} + Q) / (P_{k-1} + Q + R)$

$\beta = R / (P_{k-1} + Q + R)$

As $\alpha + \beta = 1$, we rewrote x'_k as:

$x'k = x'_{k-1} (1 - \alpha) + x_k \alpha$

The next Figure shows this simple predictor graphically.



The predictor is a function of the estimated value and the sensed value, both competing in the contribution for the prediction. As a first approach, we substituted α by the estimation error:

$\alpha = w_k = (x_k - x'_{k-1})^2 = EXO_k$

We also introduced an emotional component so that stimuli that are sensed under strong emotions are more relevant in terms of prediction. Inspiring ourselves in the delta rule, we set the learning rate to the endogenous salience. Hence, Δw_k is computed as follows:

$\Delta w_k = |END_k|(x_k - x'_{k-1})$

Even this predictor may not be as optimal as the ones it is inspired on, it provided with a good relation efficiency/adaptation that performed well over unpredictable signal dimensions.

Emotions thus close the estimation loop.

Resume

Let us recapitulate. At time step k, the emotivector has a search value s_k and is expecting x'_k, a distance of $\Delta s'_k$ and a reward of R'_k, when a sensed value x_k reaches it.

The emotivector executes as follows:

Compute error: $e_k = x_k - x'_k$

Compute exogenous. salience: $EXO_k = e^2_k$

Compute distance: $\Delta s_k = (x_k - s_k)^2$

Compute endo. salience: $END_k = \Delta s'_k - \Delta s_k$

Compute sensed reward: $R_k = \Delta s_{k-1} - \Delta s_k$

Compute sensation (if s_k exists, otherwise S sensation):

If ($R_k >= 0$ and $R_k >= R'_k$) then (S+ sensation)

If ($R_k < 0$ and $R_k < R'_k$) then ($+ sensation)

If ($R_k <= 0$ and $R_k < R'_k$) then (S- sensation)

If ($R_k < 0$ and $R_k >= R_{k'}$) then ($- sensation)

Update weight: $w_k = w_k + |END_k|e_k$

Estimate next value: $x'_{k+1} = x'_k\ (1 - w_k) + x_k\ .\ w_k$
Estimate distance: $\Delta s'_{k+1} = (x'_{k+1} - s_k)^2$
Estimate reward: $R'_{k+1} = \Delta s_k - \Delta s'_{k+1}$

Conclusions

An extension to the base agent architecture aimed at providing control mechanisms for human-like behaviour was presented, based on a salience module, that manages the emotivectors, independent mechanisms that: monitor the different dimensions of the agent perceptions; have the capability to predict their next state and; based on a model of emotion and attention, provide with information regarding both the attention focus and the sensation potential of the signal. An interesting aspect is that the signal salience is computed independently from the semantics of the signal and from the rest of the agent processing. However, an emotional exogenous control based on Damasio's somatic markers (described in Martinho, 2003) still has to be integrated, before assessing the value of the model.

7. Conclusions

In this document, we provided with an outline of the reciprocal relations between emotion and anticipation, with a strong emphasis on emotion.

We first presented the current state of the study of emotion in three fields: philosophy of mind, affective neuroscience, and psychology. We explained why philosophers usually do not study emotions. We detailed the current beliefs and directions of the research in affective neuroscience, a field that has recently made much progress. Finally, we discussed why it is difficult to reach a definition of emotion due to the plethora of existing approaches (Scherer's "emotional swamp"). However, we point a possible definition of emotion, based on the argument that this definition supports most of the current approaches to the study of emotion, be it from the philosophy of mind, affective neuroscience or psychology point of view.
We then selected a few representative architectures and application models from the field of affective computing, and presented them in some detail. From our research, it was clear that most of the current systems do not deal with the concept of anticipation explicitly, although most support planning capabilities. Anticipation is a novel approach to Affective Computing that (at the least) will provide with a valuable fresh insight on affective architectures, that appear to have somewhat stagnated in a pragmatic shallowness.

Afterwards, we presented the concept of an anticipatory system, and briefly explained its value when confronted with the universality of reactive systems. We then argued how the seminal work of Rosen in anticipatory systems shows that an affective component is needed from the start. Indeed, such systems need at least to have a constant feedback on whether the system is expected to be going in the desired or undesired direction: a function that is also attributed to emotion (not to mention surprise, an inherent characteristic of anticipatory systems). The knowledge of emotion can prove valuable in the design of such systems.

In the following section, we summarized the discussion that took place among partners, focusing on the difference between anticipatory affective systems and "normal" anticipatory systems. The discussion was presented in the context of possible test-bed scenarios for anticipatory affective architectures. From this discussion, the relation between emotion and anticipation was identified as three-fold: emotions resulting from anticipatory representations; anticipatory behaviour elicited by emotions and the meta-aspects of such a recursive process (e.g. emotions resulting from anticipated emotions). A first step towards a blueprint trying to integrate all aspects of an affective anticipatory approach was presented: the "anticipatory continuum".

Finally, two affective anticipatory approaches were presented. The first approach is a high-level more formal approach, oriented towards the integration of anticipatory-based emotions in BDI models. The second is a low-level architecture that extends the basic agent architecture, in the attempt to provide with a basic automated attention and emotion control based on sensor anticipation.

From all the presented evidence, it is clear that Anticipation and Emotion are closely related. As such, affective architectures should contemplate how they support anticipation and, conversely, anticipatory systems should contemplate how emotion is supported, as one does not exist without the other.

References

Arnold, M.B. (1970). Brain functions in emotions: a phenomenological analysis. In P. Black (Ed.) Physiological correlates of emotion. New York: Academic Press.

Averill, J. (1982). *Anger and aggression: An essay on emotion*. New York: Springer.

Bandura, A. (1982). Self-efficacy mechanism in human agency. American Psychologist, 37, 122-147.

Bates, J. (1994). The Role of Emotion in Believable Agents. In Technical Report CMU-CS-94-136, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Bechara A. (2004). The role of emotion in decision-making: evidence from neurological patients with orbitofrontal damage. Brain and Cognition, 55.

Bechara, A., Damasio H., and Damasio A.R. (2000). Emotion, decision- making and the orbitofrontal cortex. Cerebral Cortex, 10.

Burgoon, J. K. (1993). Interpersonal expectations, expectancy violations, and emotional communication. Journal of Language and Social Psychology, 12, 30-48.

Cannon, W.B. (1927). The James-Lange theory of emotion: a critical examination and an alternative theory. American Journal of Psychology.

Chalmers, D.J. (1996). The conscious mind. New York: Oxford University Press.

Damasio, A.R. (1994). Descartes' error: emotion, reason, and the human brain. New York: Avon Books, Inc.

Darwin, C.R. (1872). The expression of the motions in man and animals. London: Albermale.

Davidson, R.J. (2003). Seven sins in the study of emotion: correctives from affective neuroscience. Brain and Cognition 52. Academic Press.

Dias, J. (2005): FearNot!: Creating Emotional Autonomous Synthetic Characters for Emphatic Interactions. Universidade Técnica de Lisboa, Instituto Superior Técnico, Lisboa, MSC Thesis, 2005.

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

Elliot, C. (1992). The Affective Reasoner: A process model of emotions in a multi-agent system. Northwestern University, PhD Thesis, Illinois.

Ellsworth, P.C. (1991). Some implications of cognitive appraisal theories of emotion. In K.T. Strongman (Ed.), International review of studies of emotion, vol 1. Chichester: Wiley, pp. 143-160.

Fehr, B., Baldwin, M. (1996). Prototype and script analyses of lay people's knowledge of anger. In G. J. O. Fletcher & J. Fitness (Eds.), Knowledge structures in close relationships: A social psychological analysis (pp. 219-246). Hillsdale NJ: Erlbaum.

Firby R. (1989). Adaptative execution in complex dynamic worlds. Yale University, New Haven, CT, PhD Thesis.

Fisher, W.F., Shaver, P. and Carnochan, P. ((1990). How emotions develop and how they organize development. Cognition and Emotion, 4, pp. 81-227.

Fitness, J., Fletcher, G. J. O. (1993). Love, hate, anger, and jealousy in close relationships: A prototype and cognitive appraisal analysis. Journal of Personality and Social Psychology, 65, 942-958.

Foley, J., van Dam, Feiner, S., A., Hughes, J. (1990). Computer Graphics. Addison Wesley.
Fridja, N.H. (1986). The emotions. Cambridge: Cambridge University Press.

Gadanho S., Hallam J. (1998). Emotion triggered learning for autonomous robots. In SAB '98 Workshop on Grounding Emotions in Adaptive Systems, pages 31–36.

Gadanho S. (2003). Learning Behavior-Selection by Emotions and Cognition in Multi-Goal Robot Task. In Journal of Machine Learning Research 4, pages 385-412.

Geiss, M. C., Zwicky, A. M. (1971). On invited inferences. Linguistic Inquiry, 2, 561-566.

Gratch, J. (2000). Émile: Marshalling Passions in Training and Education. In 4th International Conference on Autonomous Agents, ACM Press, June 2000.

Gratch, J., Marsella, S. (2003). Modeling Coping Behavior in Virtual Humans: Don't Worry, Be Happy. In Proceedings of Second International Joint Conference on Autonomous Agents and Multiagent Systems, Australia.

Gray, J.A. (1987). The psychology of fear and stress. Cambridge: Cambridge University Press.

Hammond, L. (1970). Physiological Correlates of Emotion, chapter Conditioned Emotional State. Academic Press.

Harlow, H., Stagner, R. (1933). Psychology of Feelings and Emotions, chapter 2 - Theory of Emotions. Number 40. Psychological Reviews.

Hinton, G., McClelland, J. (1988). Learning representations by recirculation. In D. Anderson (Ed.) Neural Information Processing Systems. American Institute of Physics.

Izard, C.E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. Psychological review, 99, pp 561-565.

James, W. (1884). What is an emotion? Mind, 9.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. Transactions of the ASME, Journal of Basic Engineering.

Kleinginna, P.R., Kleinginna, A.M. 1981. A categorized list of emotion definitions with suggestions for a consensual definition. Motivation and Emotion, 5. pp 345—355.

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

LeDoux, J. (1996). The emotional brain: the mysterious underpinning of emotional life. New York: Touchstone.

Levitt, M. J. (1991). Attachment and close relationships: A life span perspective. In J. L. Gerwitz & W. F. Kurtines (Eds.), Intersections with attachment (pp. 183-206). Hillsdale, NJ: Erlbaum.

Lewis, M. (1993). Self-conscious emotions: embarrassment, pride, shame and guilt. In M. Lewis and J.M. Havilland (Eds), Handbook of emotions. New-York: the Guilford Press, pp. 564-574.

Maltesta-Magai, C., Izard, C.E., Camras, L. (1991). Conceptualizing early infant affect: emotions as fact, fiction or artifact?. In K.T. Strongman (Ed.), International review of studies on emotions, vol 1. Chichester: wiley, pp. 1-36.

Mandler, G. (1992). A constructivist theory of emotion. In N.S. Stein, B.L. Levental and T. Trabasso (Eds), Psychological and Biological Approaches to Emotion. Hillsdale: Erlbaum.

Marsella, S., Johnson, L., LaBore, C. (2000): Interactive Pedagogical Drama. In proceedings of the Fourth International Conference on Autonomous Agents, pp. 301-308, ACM Press, 2000.

Martinho, C. (1999). Emotions in Motion: short time development of believable pathematic agents in intelligent virtual environments. Universidade Técnica de Lisboa, Instituto Superior Técnico, Lisboa, MSC Thesis, 1999.

Martinho, C., Gomes, M., Paiva, A. (2003). Synthetic emotension: building believability. In Ayleth, R. and Rist, T. (Eds), International Workshop on Intelligent Virtual Environments. Springer-Verlag.

Mellers, B. A., Schwartz, A., Ho, K., Ritov, I. (1997). Decision affect theory: Emotional reactions to the outcomes of risky options. Psychological Science, 8, 423-429.

Miceli, M., Castelfranchi, C. (1997). Basic principles of psychic suffering: A preliminary account. Theory & Psychology, 7, 769-798.

Miceli, M., Castelfranchi, C. (2000a). The role of evaluation in cognition and social interaction. In K. Dautenhahn (Ed.), Human cognition and agent technology. Amsterdam: Benjamins.

Miceli, M., Castelfranchi, C. (2000b). Nature and mechanisms of loss of motivation. Review of General Psychology, 4, 238-263.

Miceli, M., Castelfranchi, C. (2002b). The mind and the future: The (negative) power of expectations. Theory and Psychology, 12, 335-366.

Millenson, J. (1967). Principles of Behavior Analysis. Macmillan.

Minsky,M. (1986). Society of mind. Simon and Schuster.

Muller, H., Rabbit, P. (1989) Reflexive orienting of visual attention: time course of activation and resistance to interruption. Journal of Experimental Psychology: Human Perception and Performance, (15).

Oaksford, M., Stenning, K. (1992). Reasoning with conditionals containing negated constituents. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 835-854.

Oatley, K., Jenkins, J.M. (1996). Understanding emotions. Oxford: Blackwell.

Oatley, K., Johnson-Laird, P.N. (1987). Towards a cognitive theory of emotions. Cognition and Emotion, 1.

OReilly, R., Munukata, Y. (2000). Computational explorations in cognitive neuroscience: understanding the mind by simulating the brain. MIT Press.

Ortony, A., Clore, G., Collins A. (1988). The cognitive structure of emotions. New York: Cambridge University Press.

Parrott, W. G., Schulkin, J. (1993). Neuropsychology and the cognitive nature of the emotions. Cognition and Emotion, 7, 43-59.

Picard, R. (1997). Affective Computing. MIT Press.

Plutchik, R. (1991). Emotion and evolution. In K.T. Strongman (Ed.), International review of studies of emotions, vol 1. Chichester: Wiley, pp. 37-58.

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

Parott, W.G. and Schulkin, J. (1993). Neuropsychology and the cognitive nature of the emotions. Cognition and Emotion.

Posner, M. (1980). Orienting of attention. Quaterly Journal of Experimental Psychology, (32).

Rescorla, R., Wagner, A. (1972). Classical Conditioning: II - Current Research and Theory. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. Appleton-Century-Crofts.

Rickel J., Johnson L. (1998). STEVE: A Pedagogical Agent for Virtual Reality. In Proceedings of AGENTS'98, pp.332-333.

Rivera, (1977). A structural theory of emotion. New York: International Universities Press.

Rosen, R. (1985). Anticipatory systems*: Philosophical, Mathematical and Methodological*

Foundations. New York: Pergamon Press.

Russel, S., Norvig, P. (1995). Artificial Intelligence: a Modern Approach. Prentice Hall.

Schachter, S. (1964). The interaction of cognitive and physiological determinants of emotional state. In L. Berkowitz (Ed.) Advances in experimental social psychology, Vol. 1 (, pp. 49-80). New York: Academic Press.

Scherer, K. (1984). On the nature and function of emotion: A component process approach. In Scherer K.R. and Ekman P. (Eds), Approaches to emotion. Erlbaum, Hillsdale, NJ, 293–318.

Sherman, S. J. (1980). On the self-erasing nature of errors of prediction. Journal of Personality and Social Psychology, 39, 211-221.

Sloman, A. (1998). Damasio, Descartes, alarms, and metamanagement. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC '98), pages 2652–7, San Diego, CA, USA.

Sloman, A. (2001). Beyond shallow models of emotion. Cognitive Processing, 177–198.

Smith, C.A. and Lazarus, R.S. (1990). Emotion and adaptation. In L. Pervin (Ed.), Handbook of personality: Theory and research. New York: Guilford Press.

Staller, A., Petta, P. (1998). Towards a Tractable Appraisal-Based Architecture for Situated Cognizers. In Grounding Emotions in Adaptive Systems, Workshop Notes, Fifth International Conference of the Society for Adaptative Behaviour (SAB98).

Schwarz, N. (1990). Feelings as information: Informational and motivational functions of affective states. In E.T. Higgins and R.M. Sorrentino (Eds) Handbook of motivation and cognition: Foundations of social behavior, Vol. 2 (pp. 527-561). New York: Guilford Press.

Shaver, P., Schwartz, J., Kirson, D., O'Connor, C. (1987). Emotion knowledge: Further explorations of a prototype approach. Journal of Personality and Social Psychology, 52, 1061-1086.

Thomas, F. , Johnson, O. (1994). The Illusion of Life. Hyperion Press.

Tooby, J., Cosmides, L. (1990). The past explains the present: Emotional adaptations and the structure of ancestral environment. Ethology and Sociobiology, 11, 375-424.

Velásquez, J. (1998). Modeling emotion-based decision-making. In Dolores Cañanero, editor, Proceedings of the 1998 AAAI Fall Symposium Emotional and Intelligent: The Tangled Knot of Cognition, pages 164–169, Orlando, FL, USA.

Wason, P. C., Johnson-Laird, P. N. (1972). Psychology of reasoning: Structure and content. Cambridge, MA: Harvard University Press.

Weiner, B. Russell, D., and Lerman, D. (1979). The cognition-emotion process in achievement-related contexts. Journal of Personality and Social Psychology, 37, 1211-

1220.

Wells, A., Matthews, G. (1994). Attention and Emotion - a Clinical Perspective. Psychology Press.

Widrow, B., Hoff, M. (1960). Adaptive Switching Circuits. IRE WESCON Convention Record.

Young, P. (1961). Motivation and Emotion. Wiley.

Zeelenberg, M., van Dijk, W. W., Manstead, A. S. R., van der Pligt, J. (2000). On bad decisions and disconfirmed expectancies: The psychology of regret and disappointment. Cognition and Emotion, 14, 521-541.