Machinery for artificial emotions

Luís Miguel Botelho

luis@iscte.pt

Department of Information Sciences and Technologies of ISCTE

Lisbon, Portugal

Helder Coelho

hcoelho@di.fc.ul.pt

Department of Computing Sciences

Faculty of Sciences of the University of Lisbon

Lisbon, Portugal

Contact author

Luís Miguel Botelho

Departamento de Ciências e Tecnologias da Informação

Instituto Superior de Ciências do Trabalho e da Empresa

Av. das Forças Armadas, Edifício ISCTE

1600 Lisboa, Portugal

Email: luis@iscte.pt

Fax: 351-1-7964710

Phone: 351-1-7903906

This work was partially supported by the contracts PSCH/OGE/1038/95 (Mago), 2/2.1/TIT/1662/95 (Sara) and by UNIDE/ISCTE.

Abstract

We present a preliminary definition and a theory of emotion viewed as a sequential process comprising the appraisal of the agent global state, the generation of an emotion-signal and an emotion-response. This theory distinguishes cognitive from affective appraisal on an architecture-grounded basis. A scheme for emotion classification with five dimensions is presented. Among them, we emphasize the roles played by emotions and the way these roles are fulfilled. It is shown how emotions are generated, represented and used in the Salt & Pepper architecture for autonomous agents [Botelho 1997]: emotion-signals, represented by label, valence and intensity, are generated by the affective engine through the appraisal of the agent's global state; for each emotion-signal there are several nodes stored and interconnected in the behavioral or in the cognitive networks of the agent; each of these nodes contains an emotion response that may be executed when an emotion-signal is generated; emotion intensity relates to the activation of the node. The implementation of a concrete example is fully described.

1. Introduction

In this paper we discuss emotion in the perspective of the Salt & Pepper Project. The aim is to develop and evaluate computer agent architectures for general intelligence. The approach taken in the Salt & Pepper Project is to build mechanisms and features that can play the same roles for artificial agents as those responsible for the success of natural agents, notably human beings. In particular we have taken inspiration in cognitive science and in neuroscience.

Emotion has been identified as one of the key elements of the intelligence and of the adaptive nature of human beings [Damásio 1994][Goleman 1995][Bechara et al 1997], therefore we want to develop artificial mechanisms that can play the role emotion plays in natural life. We refer these sort of mechanisms by "artificial emotions".

Natural emotions are the key to the successful use of “pure” problem solving and decision making skills as measured by intelligence tests like the IQ and the SAT. Emotions play a role at a meta-level of control: they help the organism to regulate itself in order to improve its adaptation to the environment. Artificial emotions will do the same. They may lead to reflexive behavior, they may intensify the motivation of an agent, they can create new motivators (e.g., new goals), they can set new criteria for the selection of the methods and the plans the agent uses to satisfy its motives, and they provide the basis for adaptive learning.

Since we are dealing with artificial emotions, the psychologically inclined reader should not be surprised to find out that this article addresses emotions that do not forcefully match the usual natural emotions such as joy and relief. Instead, we focus on lower level emotions such as performance evaluators and attention-shift warnings.

A performance evaluator is an emotion-signal that evaluates the performance of the agent in a given task. It is generated by the agent to warn it that it has been performing poorly or to inform it that it has been performing exceedingly well. Like all other emotions, performance evaluators have an adaptive use. For instance, consider a young man trying to be a professional piano player that always blushes and trembles in front of his audiences. This might be an emotion of performance evaluation warning the young man that he should consider another profession or that he should rehearse better before performing in public.

Attention shift warnings signal the agent that it should pay attention to something else. Think of a man walking across a street fully absorbed by his thoughts that suddenly jumps backwards avoiding being hit by a bicycle he has not noticed just until that moment. This is a case in which an emotion signal of attention-shift warned the man that he should be paying attention to what had been going on in the street, instead of being thinking of something else. As a response to this emotion-signal, the man attended the traffic in the street and managed to avoid being hit.

In both of these examples, emotion plays an important role at the control level of human behavior. Our research goal is to create mechanisms with a similar role for computational agents.

In the field of Artificial Autonomous Agents, the mainstream research on emotion is driven by the believable agents community. In that community, emotion is viewed as a way of improving the believability of artificial agents. Joseph Bates and the Oz project [Bates et al 1992] are specially concerned with interactive drama characters. Emotion has also been used as a way to improve interface agents and, more recently, avatars [Vilhjálmsson and Cassell 1998]. From the point of view of this community emotion is studied as a vehicle to create agents with more natural expressive behavior. Emotional expressive behavior is an end in itself.

Clark Elliott and the Affective Reasoning Project [Elliott 1994, 1998] as well as the Kimsac Project [Arafa, Charlton and Mamdani 1998] aim at improving the agent's performance. They have shown that emotion can improve the way information is created, conveyed and understood. This research goal can also be found in other work such as [Numao et al 1997] in which emotion is used to improve the music-arrangement capabilities of an intelligent system. From this point of view, emotion is a vehicle to achieve better performance, but it is addressed only at the task level. This work does not focus on the relation between affective phenomena like emotion and artificial agents architectures.

Some other researchers such as Dolores Cañamero [Cañamero 1997], Luís Botelho and Helder Coelho [Botelho and Coelho 1996b, 1998a, 1998b] and Aaron Sloman and the Cognition and Affect Project [Sloman 1987][Beaudoin and Sloman 1993] have argued that emotion should be considered at the level of the agent architecture.

Gordon Bower [Bower 1981] presents a memory-based cognitive theory of emotion that explains mood congruence and state dependence effects, and the generation of autonomic behaviors in response to the activation of emotions, but he doesn’t present any mechanism for emotion elicitation.

The work of Andrew Ortony and his group [Ortony et al 1988] laid down the basis to a widely accepted classification of human emotions together with their eliciting conditions.

Sloman and the Cognition and Affect Group [Sloman 1987] have been studying classes of architectures in which both cognitive and affective phenomena emerge as the result of essentially the same kind of processing. These architectures provide a deeper understanding of the relationship between emotion and behavior. A major contribution of this work is the idea that the exact emotion response to a particular situation depends on the motivators generated on the whole process. This means that the same situation may trigger different emotion responses.

[Botelho 1997] describes a concrete architecture for autonomous agents (Salt & Pepper) and shows that moods and emotions play important roles in the agent’s behavior through its impact on cognitive processes such as the allocation of cognitive resources [Botelho and Coelho 1996a], attention control [Botelho and Coelho 1996b] and adaptive learning [Botelho and Coelho 1998b]. However, moods and emotions are taken for granted and little is said about their generation besides some suggestions about the theory and technology to be used.

The present article presents a preliminary theory of artificial emotion and describes a particular architecture for artificial autonomous agents that integrates all the above contributions in a single unified framework. This framework describes the emotion process from the appraisal stage until the behavioral response. It explains mood congruence, state dependence, the generation of autonomic behaviors in response to the activation of emotion, it explains why the same external situation may lead to different emotion responses and it explains how several motivators may be generated during the emotion process. It also explains that the emotion response may have several components appearing simultaneously or in sequence. Finally, in this framework, the emotion process is an adaptive process that improves along the agent's life.

This article is built upon two previous papers by the authors. The first presented at the second International Conference on Autonomous Agents [Botelho and Coelho 1998a] and the second presented at the workshop Grounding Emotions in Adaptive Systems of the fifth International Conference on Simulation of Adaptive Behavior [Botelho and Coelho 1998b].

Our main contributions are (i) the introduction of a process view of emotions; (ii) an original proposal concerning the classification of emotions; and (iii) a concrete implementation and use of emotion within Salt & Pepper architecture [Botelho 1997].

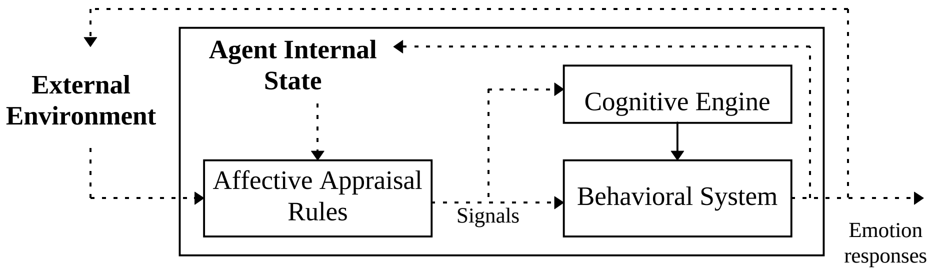
In section 2, we present our ideas and work regarding emotions, starting by a "preliminary definition" of emotion. In the first subsection, we propose a matrix classification of emotions with five classification dimensions: the role/function of emotions, the mediating process by which emotions fulfill their roles, the object of appraisal, the urgency to initiate repairing processes, and the type of appraisal that elicits emotions. Afterwards, we describe our theoretical ideas about the emotion generation process, from appraisal to emotion response. In the third subsection, we discuss the appraisal process. Section 2 proceeds with some background information regarding the Salt & Pepper architecture. Then, we show how the theoretical ideas presented are implemented and used within the Salt & Pepper architecture. In the last subsection, we described an implemented example of the appraisal process. In section 3, we compare our approach with related work and present some conclusions and several issues for future research.

2. Emotions: from theory to implementation

Although we are not ready yet to provide an accurate definition of emotion, we will present our own view, hoping it will constrain the use of the concept in sensible and useful ways.

Emotion is a process that involves appraisal stages, generation of signals used to regulate the agent's behavior, and emotion responses. In the first stage, a set of appraisal structures is used to evaluate the global state of the agent (internal state plus external environment). If certain conditions hold, an emotion-signal is generated informing the agent of the result of the appraisal stage. A signal can have positive or negative valence. If the agent global state is found to conflict with the motives of the agent (e.g., instincts, goals, intentions, values, attitudes), a negative emotion-signal is generated. If the agent global state is specially favorable according to the agent motives, a negative signal is generated. The quality of the signal is related to the behavioral response that will be produced. Signals generated during an appraisal stage are sent to the behavioral system of the agent possibly giving rise to a behavioral response – the emotion response. In different situations, the same emotion-signal may give rise to different emotion-responses, depending on the internal state of the agent. Among other things, an emotion-response may be a reflexive (overt or covert) action, it may increase or decrease the resources assigned to satisfy a given motive, it create a new motivator (e.g., a new goal), it may set a new criteria to be used for selecting plans (or other methods) to achieve current motives. After an emotion-signal has been generated and an emotion-response has been performed, the global state of the agent changes and a new appraisal is done possibly generating a new emotion-signal. As sketched in Figure 1, this continuous process mirrors the concept of "feedback loop" of the control theory. The emotion process goes on in parallel with the cognitive process of the agent.

A simple metaphor for our model is an operating system, where the exceptions play the role of the emotion-signals, the actions performed by the exception-handlers play the role of the emotion-responses and the appraisal stage is played by all if-statements in the operating system code that generate exceptions.



Only the emotion process and the components of the agent architecture relevant to it are depicted. Other interactions involving the cognitive engine are not shown. The process by which emotion is learned is also omitted from the figure.

Figure 1 - Emotion process

Although motivation and emotion are both control processes directed at the satisfaction of the agent motives, there is an important difference between them (but see also [Aubé 1998]). In the emotion process, there is no direct coupling between the signals generated by the appraisal stage and the behaviors enacted in response to those signals. In the motivation process, the cognitive system of the agent analyzes the agent global state and selects the appropriate behavioral response (e.g., it makes decisions, builds plans and selects actions). One could argue that this distinction may be related to the distinction between automatic (non conscious) and deliberative (conscious) processing. The feeling of an emotion (which makes the agent aware of it and hence conscious) would be a result of the way signals are generated in specific architectures (e.g., the release of certain transmitters) or a result of the emotion-response (e.g., muscle contraction).

We believe this "preliminary definition" preserves much of the common sense understanding of emotion and identifies precise features that enable us to distinguish it from other concepts.

Among other things, emotion can be used as a basis for attention control, for performance evaluation and regulation, for identification of and recovery from malfunctioning-components, and for adaptive learning processes.

Matrix classification of emotions

Emotions have been classified in a number of ways. Sometimes a different classification is proposed because a different feature is being classified. Herein we show that different classification proposals address different components of the global emotion process and we present a classification scheme based on our definition of emotion.

Ritter [Ritter 1993] proposes two broad classes: dry and wet emotions, according to the processes involved in their generation. Dry emotions aren't really felt by the organism, they exist only upon cognitive reflection about the current state of affairs by an observer (external or internal). Wet emotions are due to changes of the information processing architecture of the organism (e.g., resource depletion that may affect the available cognitive operators). According to our process view of emotion, the classification proposed by Ritter addresses the appraisal process.

[Ortony et al 1988] proposes a classification of emotions based on the stimulus causing the emotion: those induced by events, those induced by agents, and those induced by objects. Emotions result of the appraisal of the external stimulus according to the agent's goals, standards and attitudes. For instance, if an event impairs the achievement of one of the agent's goals, a negative valence emotion is generated, if an observed agent behaves according to one of the observer's standards, a positive emotion is generated. From these basic categories, [Ortony et al 1988] presents a complex taxonomy of emotions that is probably the most widely used in AI, namely in the Oz and the Affective Reasoning projects.

According to our process view of emotion, the classification scheme proposed by Ortony's group addresses only the appraisal process.

Sloman [Sloman 1987][Sloman 1995] distinguishes two main groups of emotions: some emotions are signals that the behavior must be changed because there is a conflict between the agent's goals and current state and plans; other emotions are "states" of constant interruption of the agent's current processing by insistent motivators, making it hard for the agent to concentrate on higher priority tasks. According to our model, the first of these categories refers the emotion-signals, whereas the second category refers the effects of the emotion-response. Actually, an emotion-signal of attention-shift may cause the current cognitive task of the agent to be interrupted (i.e., replaced by another one).

We present an emotion classification scheme based on our stepwise view of emotion as a sequential iterative process. Our classification scheme accommodates the distinctions pointed out by the referred authors. We propose five dimensions for the classification of emotions (Table 1), each of which can be further decomposed into a class-subclass taxonomy.

Dimension of classification	**Examples**	**Process component**
Role/function of emotion	Attention shift warning, performance evaluation, malfunctioning-component warning, motivation intensifier	Emotion-signal
Process by which emotion fulfills its role	Reflexive action, creation of motivators, setting plan selection criteria	Emotion-response
Urgency of the repairing process	Urgent (e.g., need to immediately attend the external environment), not urgent (e.g., need for long-term improvement of default criteria for plan selection)	Emotion-response
Object of appraisal	External environment, internal state, past events, current events	Appraisal stage
Type of appraisal	Affective appraisal, cognitive appraisal	Appraisal stage

Table 1 - Dimensions of emotion classification

Table 1 and the explanations that follow are independent of the architecture used and the context in which the actual emotion takes place. Therefore, most details (e.g., "whens" and "hows") are left unspecified. In the last two subsections of this section, we will discuss some of these concepts in the framework of the Salt & Pepper architecture for autonomous agents.

Here, we focus on two kinds of emotions regarding their role in autonomous agents architectures: attention shift warnings and performance evaluators. Upon the generation of an attention shift warning, the agent's information processing should consider the possibility of interrupting the current cognitive task and direct the agent's attention to something else (e.g., external stimuli, internal goals). Whether and how attention really shifts depend on the specific architecture and maybe on the current context in which the warning takes place. When an emotion of performance evaluation is generated, the agent's information processing should consider regulating its behavior according to the nature of the evaluation. Again, the question of when and how this happens depends on the architecture and on the circumstances.

Besides attention shift warnings and performance evaluation signals, emotions can play other roles, such as motivation intensification and malfunctioning-component detection.

The emergence of emotions is associated with emotion responses. The emotion response should have an adaptive nature, that is the agent should behave in a way that is adequate to the meaning of the emotion. For instance, if an attention shift warning is generated, the agent may react by reflexively directing its sensors to some aspect of the external environment. This is the mediating process by which an emotion fulfills its role. Instead of a reflexive action, the emotion response might be the generation of the goal of attending the external environment in ten seconds or so. In many architectures, this will involve some kind of deliberative process. Besides reflexes and goals, the information processing of the agent might set appropriate criteria for plan selection. For instance, the agent might select the plan that is believed to *satisfice* one of its goals and to contain planned actions for monitoring the external environment. All the mediating mechanisms proposed in this paper are consonant with the view that emotions operate at a meta level of control [Goleman 1995]. It is also worth comparing the kind of affective learning that people acquire while they get older (decreasing the frequency of emotional interruption of current plans) [Simon 1967] with the capability of selecting plans that minimize the likelihood of generation of attention shift warnings.

An adequate emotion response also depends on what has caused it. At least, the agent needs to "know" [a] whether the emotion was generated due to the external environment or to its internal state. If an agent detects an attention shift warning originated in the external environment, its attention should be directed to the outside world. If the origin of the warning is the agent's internal state, then the agent has to attend internal state variables (e.g., attend an important goal).

The agent's information-processing mechanism should also "know" how much time i s there available before any behavioral response is undertaken. This is the urgency of the repairing process. In case of an urgent warning, the emotion response should be fast, probably a reflexive action. In case of non-urgent warnings, there is no need to immediate action: the creation of a new goal is probably an adequate emotion response.

Finally, emotions follow some form of appraisal. According to mainstream research, this is a cognitive process. In the next subsection we argue that the appraisal process that mediates the generation of an emotion may be either a cognitive or an affective process, even in architectures within the information-processing paradigm. In fact, this is an important distinction, specially if one is considering agents of limited rationality acting in dynamic challenging environments: a cognitive appraisal consumes resources that might be needed for an important and urgent deliberative process or problem solving whereas an affective appraisal *per se* won't disturb the agent's performance on cognitive tasks. Besides this issue about resources, it is likely that a cognitive appraisal gives rise to an emotion that differs, in some way, from those originated by an affective appraisal, but we won't further explore this possibility in this paper.

Now, we concentrate on the first two dimensions of emotion classification in more detail: the role/function of emotions and the mediating process by which emotions fulfill their roles. Table 2 describes some possible general emotion responses relating these two dimensions of emotion classification. Numbers in parentheses identify each of the six cases considered. Text in parentheses indicates an optional component of the emotion response. Cases (4.a) and (4.b) of Table 2 represent alternative emotion responses to a reflexive/autonomic performance evaluator emotion.

	Attention shift warnings	**Performance evaluators**
Reflexive/autonomic	(1) Interrupt current processing; initiate new processing.	(4.a) Trigger autonomic adaptive learning processes that avoid undesirable outcomes or achieve desirable states of affairs. (4.b) Perform an action that repairs undesirable outcomes due to poor performance, or maintain desirable states of affairs
Motivator creation	(2) Interrupt current processing; create delayed attention shift goal; (resume current processing)	(5) Generate goals whose fulfillment repairs undesirable outcomes due to poor performance, or maintains desirable states of affairs
Plan selection	(3) Adopt plans with regular monitoring of the external environment	(6) Adopt plans that contain actions that repair undesirable outcomes, or maintain a desirable state of affairs

Table 2 - Role/function of emotion vs. Mediating process

Attention shift warnings (column 1 of Table 2) are useful when an external event occurs that represents an important opportunity or risk for the agent. Case (1) of Table 2 can be illustrated with several examples. For instance, consider an agent processing stock yields available in the Internet. Suppose the agent is currently planning to move to another stock information server because it detects a very low transmission rate. Suddenly, a particular stock starts to drop very quickly. This sudden change might have been unnoticed if an attention shift warning wouldn't have directed the agent's attention to the stock information, interrupting its current plan for finding another server.

In case (2) an agent is processing information packets delivered through some communication channel. Since the average delivery rate is very high compared to the average information packet processing time, while the agent processes a packet, other packets get lost. Suppose that while the agent is processing packet p_1, packets p_2, p_3 and p_4 would be lost. Suppose also that packet p_3 contains very important information for the agent. The agent would have missed it if the emotion response to an attention shift warning wouldn't have caught p_3, interrupted the processing of p_1, created the goal "process p_3" and resumed the processing of p_1.

In case (3), consider the same agent of case (2). This time, the emotion response to the attention shift warning doesn't create the goal "process p_3". Instead, a new plan selection criteria is defined. For instance, a plan can only be selected if the processing of information packets is intertwined with actions to monitor new information packets and schedule important packets for subsequent processing.

In case (4), suppose the agent's information processing is stuck in a endless loop. The agent would still be stuck, if the emotional reflexive response to the generated performance evaluator wouldn't have exited that loop.

In case (5), consider an agent that tries to get two files from the same host, using ftp. As the transmission rate while getting the first file is very low, an emotion of performance evaluation is generated that replaces the goal of getting the second file immediately after getting the first, by the new goal of getting it during the night.

In case (6) the emotion response of the agent is at the plan-selection level. If an agent fails to complete the classification of a sequence of video frames, a negative performance evaluator is generated. Upon detecting that emotion, the agent defines a new criteria for selecting multimedia analysis algorithms.

Affective vs. cognitive appraisal

Given a particular architecture, it should be possible (at least in principle) to identify those modules responsible for the affective components and those clearly on the side of the cognitive components of the whole information processing (from now on, Affective Engine and Cognitive Engine, respectively). According to recent literature [Damásio 1994][Goleman 1995][Balkenius and Morén 1998], this kind of distinction has also been found in the human brain. For instance, whereas the amygdala is thought to be responsible for affective phenomena, the visual cortex is thought to be on the side of cognition.

Even if we are working within the information-processing paradigm (architectures implemented by computers), if the appraisal is performed by some part of the Cognitive Engine, that's a cognitive appraisal, if it is made by some part of the Affective Engine, we are free to talk about affective appraisal. After all, although the operation of the limbic system can be modeled within the information processing paradigm, we still say it is responsible for affective phenomena. However, this distinction between affective appraisal and cognitive appraisal only makes sense if it is not merely based on fancy but meaningless designations we give to the modules of our architectures. What then is the essential difference between the Affective and the Cognitive engines?

As repeatedly told by Aaron Sloman, [Sloman 1987][Sloman 1995] the nature of both cognitive and affective processing is essentially the same. Nevertheless, there are at least three aspects that should differentiate the Affective from the Cognitive engines in an architecture for autonomous agents: the kind of information processed in each of those subsystems, the purposes of such processing (what do we need it for?), and the typical reaction time.

kind of information processed

In general terms, the Affective Engine processes information concerning the hypothetical or the actual satisfiability of the agent's motives (e.g., needs, goals, values, standards), whereas the Cognitive Engine may process all this information plus problem solving information, decision making information and declarative information describing several aspects of the world (including the agent itself).

purpose of information processing

The main purpose of information processing of the Affective Engine is to generate signals that may be used to help the Cognitive Engine perform its tasks. Those signals can help to select cognitive structures more likely to be relevant in each context, they can help to control the focus of attention of the Cognitive Engine, they can help to determine how much effort is spent by the agent in any given task, they can help deciding what goals to choose, what kinds of plans to use, and what alternative outcomes to prefer. The main tasks of the Cognitive Engine are goal satisfaction (e.g., through planning), problem solving (e.g., through heuristic searching) and decision making (e.g., through rule-based reasoning).

typical reaction-time

The Affective Engine should react much faster than the Cognitive Engine, or else it wouldn't have much use. How can this be achieved? The Affective Engine should only operate on a small fraction of the total information required by the Cognitive Engine (both external and internal information); the outcomes of the Affective Engine should be based on a smaller set of conditions than those of the Cognitive Engine; the affective processing chains should be much smaller than the cognitive ones; there may be short cuts from the Affective Engine to the output channels, and from the input channels to the Affective Engine. The kind of technology and organization of the Affective Engine should also contribute to faster decisions.

Affective Appraisal

According to [Srull and Wyer 1986], [Sloman 1987], [Ortony et al 1988], [Bates et al 1992] and [Elliott 1994], the generation of emotions depends on the motives whose fulfillment is facilitated or impaired, on the causes and instruments of that facilitation or impairment, on the importance of those motives, and on the effort already spent to achieve them. Our proposal is very much in the same line, but we stress an important difference.

Since emotions have an adaptive role, at least some emotion responses should be triggered very quickly. This means the affective appraisal should be very fast. Therefore, we propose it should avoid explicit comparisons between the current global state and the motives of the agent. The motives facilitated or impaired by the current global state should be implicit in simple domain-specific productions. Figure 2 illustrates the idea with a naive example of a comparison with the survival instinct.

Explicit comparison with the survival instinct (Figure 2, a)
if someone risks dying, he or she will feel a lot of fear; risks_dying(A, Risk) $\Rightarrow$ **activate**(fear, negative, Risk) *if someone risks running out of food, he or she risks dying;* risks_running_out_of_food(A, Risk) $\Rightarrow$ risks_dying(A, f_1(Risk)) *if someone risks running out of money, he or she risks running out of food;* risks_running_out_of_money(A, Risk) $\Rightarrow$ risks_running_out_of_food(A, f_2(Risk)) *if someone loses some amount of money, he or she risks running out of money* loses_money(A, Loss) $\Rightarrow$ risks_running_out_of_money(A, f_3(Loss))
Implicit comparison with the survival instinct (Figure 2, b)
loses_money(A, Loss) $\Rightarrow$ **activate**(fear, negative, $f_1(f_2(f_3(\text{Loss})))$)

Figure 2 - Appraisal rules

The production in Figure 2(b) detects a situation in which the survival instinct is implicitly threatened. The rationale for that production is the chain of productions represented in Figure 2(a).

As described in [Botelho and Coelho 1998b], the component of the Affective Engine responsible for the affective appraisal continuously compiles general explicit long comparison chains into small domain-specific rules, using a technique similar to the SOAR "chunking" [Laird et al 1986]. The general and the specific rules are organized in a hierarchy in which the longer chains are used only in case there isn't any applicable specific production available. Besides this hierarchical organization along the general-specific dimension, specific rules should be indexed by the domain, in order to reduce search. Finally, the affective appraisal of a situation should be limited in time.

Background of Salt & Pepper

The Salt & Pepper architecture for autonomous agents (Figure 3) has three major blocks: the Cognitive Engine, the Affective Engine and the Interrupt Manager. The Affective Engine includes affective sensors, the Affective Generator and the Affective Monitor. These two operate together to initiate the process of emotion generation. All other modules, except the Interrupt Manager, comprise the Cognitive Engine. All components of the architecture operate in parallel with and largely independently from each other.

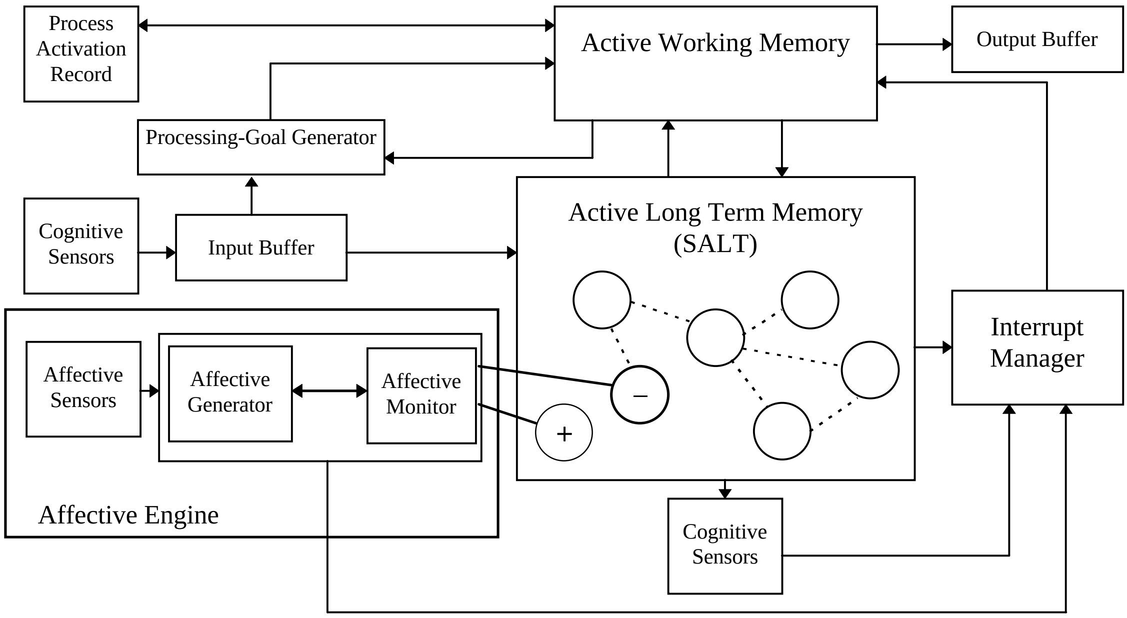


Figure 3 - Salt & Pepper Architecture for Autonomous Agents

Since the Cognitive Engine and the Affective Engine are mutually independent and differ with respect to the kind of information processed, the purpose of such processing, and the typical reaction-time, we are free to say that an appraisal performed by the Affective Generator is an affective appraisal and one that is performed in working memory is a cognitive appraisal.

long term memory

Long term memory is an associative network with spreading activation [Botelho and Coelho 1995]. Each node of the network has an identification, an activation level, a set of associations to other nodes, a header, and contains a set of related symbolic cognitive structures representing motives, declarative knowledge, plans and actions. The activation level of a node is a measure of its accessibility. When long term memory is searched, more activated nodes are found first than less activated ones. The activation of any node decays exponentially with time. The header of a node is a summary of its contents, and is used in pattern-matching operations when information is searched in long term memory. The activation received by a node spreads to other nodes through the associations between them. Stronger associations are traversed by greater proportions of activation.

control

The Input Buffer and the Affective Generator activate nodes in long-term memory (due to the presence of new external stimuli, or to emotion eliciting conditions). The Cognitive and the Affective monitors propose certain nodes to gain the agent's attention. Whenever a node is proposed to gain the agent's attention, the Interrupt Manager decides whether or not to interrupt the current cognitive processing and load the contents of the node in working memory to be processed. When the contents of a node are processed in Working Memory, the node receives a certain amount of activation, becoming more accessible.

episodic memory

Each time a decision is made, a problem is solved, or some information is acquired, in response to some interaction, the information processing mechanism of the agent creates a node in long term memory describing the episode. These nodes are called episodic nodes and constitute the episodic memory. Episodic nodes specify when the interaction took place, what was the problem that the agent had to solve, what nodes were used to solve it, what was the solution or decision outcome found, what was the source of the information acquired. In short, episodic memory contains a trace of the interactions of the agent.

context dependent behavior

Salt & Pepper has the important property of context dependent behavior. The history of the interactions of the agent determines the accessibility of nodes in long term memory and the strength of the associations between them. Therefore, if the agent is given “a proper education”, it is very likely that, in each context, more adequate cognitive structures are more accessible in long term memory. This means the behavior of the agent will be appropriate to each context.

Emotion flow in Salt & Pepper

In this subsection we show how the theoretical issues about emotion discussed along section 2 are implemented and used in the Salt & Pepper architecture.

A long term memory node is a data structure (N) with (i) the node identification (Id); (ii) the node activation level (L); (iii) the header of the node (H); (iv) an internal database containing a set of related cognitive structures (Δ); and (v) the node’s associations (A), that is, a possible empty set of pairs <Id, S> in which Id is the identification of a node to which N is associated and S is the strength of the association. Thus, a node is described by a five-tuple <Id, L, H, Δ, A>.

In the present framework, emotions are described by a set of parameters: (i) label (E) representing the emotion class and a list of instantiated arguments such as the object of the appraisal (e.g., *attention_shift_warning(object:external_environment)*); (ii) valence (V) that may have the values *positive*, *negative* and *neutral*; (iii) intensity (I); (iv) emotion program (P) representing a possibly null sequence of actions that are performed immediately upon the appraisal stage has generated an emotion-signal; and (v) emotion response (R) that is performed only if a long term memory node matching the label of the emotion-signal is selected and processed in working memory. Thus, an emotion is represented by a five-tuple

<E,V,I,P,R>.

An emotion program is different from an emotion response because the former is executed by the Affective Generator without interrupting the agent's current cognitive processing (and, arguably, without its conscious awareness) whereas the latter is performed in the agent's working memory. The execution of an emotion program could be seen as an autonomic response. In general, the emotion program creates some temporary data structures required by the emotion response. For instance, if an emotion response is the creation of a goal to process a certain message that otherwise would be lost by the agent, the attention-shift warning generated by the affective appraisal stage has really to catch and store the message so that the agent can process it afterwards (see the discussion of case (2) of Table 2 in the first subsection of section 2). The emotion-program in the Salt & Pepper architecture does this kind of job.

In the Salt & Pepper architecture, the Affective Generator performs a partial evaluation of the global state of the agent (external and internal) -- the affective appraisal. If the eliciting conditions of a particular emotion are met, the Affective Generator generates an emotion-signal (label, intensity and valence) and executes the emotion program. The Affective Monitor seeks its fragment of long term memory until it finds the most accessible node whose header matches the emotion label and has the same valence of the generated emotion-signal. This node is activated with an activation level that is a function of the generated emotion intensity (emotion activation function).

Consider an agent with a set of emotions with labels {A, B, C, D} and a set of long term nodes {A_1, A_2, B_1, B_2, B_3, C, D} representing emotions.

Node	Valence	**Emotion Label** (specified in the header of the node)	**Emotion response (Δ)**	**Activation Level** (accessibility)
A_1	positive	A	RA_1	20
A_2	negative	A	RA_2	15
B_1	positive	B	RB_1	20
B_2	negative	B	RB_2	10
B_3	negative	B	RB_3	5
C	neutral	C	RC	8
D	positive	D	RD	8

Figure 4 - Long term representations of emotions

Suppose, the Affective Generator generates the an emotion-signal with label B, negative valence and intensity 15. The node B_2 is the most accessible representation that matches the label B and has negative valence, therefore the activation of B_2 becomes 25 units (assuming an emotion activation function f(x)=x). Notice that nodes A_1, A_2 and B_1 are more activated than B_2, but neither A_1 nor A_2 matches the label B, and the valence of B_1 is positive whereas the valence of the generated emotion is negative. The header of B_3 matches the emotion label B, but B_3 is less accessible than B_2 ($B_2.L>B_3.L$), therefore B_2 is found before B_3.

Assuming B_2 becomes more accessible than the node guiding the current cognitive processing in the agent's working memory, the Affective Monitor proposes it to gain the agent's attention. If the Interrupt Manager allows, B_2 is selected to and processed in working memory. Therefore, response RB_2 will be executed. Maybe RB_2 is a reflexive action like "skip backwards", or maybe the action "create a goal of being home in about 30 minutes", for which some form of deliberative planing would be required.

The described process could have gone different ways. For instance, after the emotion label B has been generated, B_1 could have become more activated than B_2, therefore B_2 wouldn't have been selected to working memory. It could also happen that after B_2 receives 15 activation units due to the generation of the emotion-signal with label B, it doesn't become activated enough to be selected to working memory. In this case, the emotion response wouldn't be executed. Nevertheless, the activation of B_2 would still bias subsequent information processing since B_2 and the nodes to which it is associated would become more accessible in long term memory, increasing the likelihood of their future selection to working memory. This explains the congruency effect studied in psychology [Forgas and Bower 1987][Mayer et al 1992].

The above description makes it clear that the exact response of the agent to a partial evaluation of the global state depends on several factors, such as (i) the pattern of activation of all nodes in long term memory; (ii) the responses specified in the nodes that represent emotions; and (iii) the available plans that can be used to satisfy possibly created motivators.

This context-dependent mechanism is enough to explain Sloman' s proposal [Sloman 1987] according to which the same evaluation can generate hate towards someone, but it can also generate disappointment with someone, depending on the motivators that are created in the way. According to Sloman, hate will appear only if retributive motivators are generated. In our framework, the motivators involved in a particular instance of an emotion result of the execution of the program specified in the node selected to working memory in response to the signal generated in the affective appraisal stage.

The next subsection gives an example of the affective appraisal process, since the cognitive component of the architecture and its attention shift mechanisms have been explained elsewhere [Botelho and Coelho 1995][Botelho and Coelho 1996a][Botelho and Coelho 1996b][Botelho and Coelho 1996c][Botelho 1997].

The case of attention shift warnings

Here, we analyze an example that illustrates the case (1) of Table 2, in which emotions serve as attention shift warnings.

In this example, an agent picks information packets (e.g., sentences) from a communication channel, analyzes their content and acts accordingly. Sometimes the information packets are delivered by the communication channel at such a rate that the agent is incapable of completing the analysis of one packet before the next is delivered, therefore some packets get lost. In situations in which recent events are more relevant to the agent' s performance than old events, loosing recent information packets impairs the satisfaction of the motives of the agent. We are assuming that the agent is built with the implicit motive of avoiding missing relevant information (this can be seen as an instinct). Therefore, the affective appraisal mechanism of the agent should be design in such a way as to prevent the agent from missing important information packets.

% Rule 1

% Update instant of the completion of the analysis of the last
% analyzed packet

```
instant_of_last_analysis(T1),
analysis_outcome_event(Event),
event_time(Event, T2),
T2 @> T1
   => update_instant_of_last_analysis(T2),
      reset(warning_level),
      reset(lost_packets).
```

% Rule 2

```
% New packet delivered
new_packet(on)
   => set(new_packet(off)),
      allow_warnings,
      increase_lost_packets.
```

% Rule 3

```
% Number of lost packets greater than maximum allowed
lost_packets(N),
maximum_allowed(Max),
N > Max,
warnings_allowed,
warning_level(Level)
   => disallow_warnings,
      NewLevel is Level + 5,
      set(warning_level(NewLevel)),
      activate(attention_shift_warning, negative, NewLevel).
```

Figure 5 - Example of affective appraisal productions

Accordingly, it might be useful to have a mechanism that prevents the agent from spending much time with an old information packet and ignoring new more relevant packets. Yet, in time intervals in which the delivery rate is slow, the agent should be allowed to proceed its analysis of the current information packet. This means that a time-out mechanism is not adequate. What is needed, in this simple case, is to generate a warning that the agent should stop processing the current packet and attend the next packet if more than a pre specified (or otherwise learned) number of information packets were lost. For the sake of simplicity, we dispense with input buffering.

With the appropriate initial conditions the production system of Figure 5 will do the required job.

The affective engine monitors the delivery of information packets to the agent and the completion of packet-analysis episodes. When a new packet is delivered to the agent through the communication channel, the flag *new_packet* is set to *ON*. When the analysis of a packet is completed, a new assertion of the predicate *analysis_outcome_event/1* is stored in the episodic memory of the agent representing an information-packet analysis event.

Rule 2 of the affective production system of Figure 5 monitors the delivery of information packets to the agent. Whenever a new packet is received, Rule 2 increases the counter of lost packets (action *increase_lost_packets*) and resets the *new_packet* flag. Whenever a new packet is analyzed, Rule 1 resets the counter of lost packets and updates the instant of time of the completion of the last packet-analysis episode.

Rule 3 monitors the number of lost packets. When this number exceeds a certain limit, it generates an attention shift warning. This emotion-signal is sent to the behavioral-system activating a behavioral node that matches a negative attention shift warning. The selected behavioral node may contain a reflexive action like the one depicted in Figure 6.

The warning intensity is increased whenever the emotion-signal is generated to ensure that the node with the emotion-response will quickly become more activated than the node guiding the current cognitive processing in working memory. If the behavioral node containing the emotion-response is activated but it doesn't gain the agent's attention, next time another packet is lost, it will become much more activated increasing the odds the emotion response is performed.

In the affective appraisal process described, the affective sensors just need to detect the delivery of a new information packet, which is not much to ask. Since the Affective Generator must know when a new packet-analysis is finished, the Affective Monitor must also have access to the episodic nodes of long term memory. This design constraint doesn't apply only to the particular case described; in general, the affective appraisal needs to have information regarding what the agent has done or decided, that is, information kept by episodic nodes.

Figure 6 depicts a possible emotion response to an attention shift warning specified in one of the nodes representing it. When executed, it directs the agent' s attention to the communication channel and process the next information packet.

```
procedure attention_shift_warning()
begin
   type_packet packet;
   disable_interrupts();
   read_input_packet(packet);
   enable_interrupts();
   process_packet(packet);
end
```

Figure 6 - Reflexive response to an attention shift warning

In the previous example, the attention shift warning was due to an evaluation based solely on the number of information packets lost since the last analyzed packet. However, this might be an over simplistic policy. It might be desirable that an attention shift warning is generated when the communication channel delivers an information packet with certain contents. In this case, besides (or instead of) counting the number of lost packets, the Affective Generator should perform a simple content analysis of the information packets. If the contents of any packet are considered relevant, the Affective Generator should generate an attention shift warning.

In the described way, it would seem that more appraisal dimensions could be indefinitely added to the Affective Generator, however care must be taken not to overload it because the processing time required will become prohibitive. If the Affective Generator really has to consider many different aspects of the global state, it is advisable to consider a distributed solution. That is, instead of a monolithic architecture, the Affective Generator should be made of several specialized modules plugged in parallel, each of which dedicated to a specific aspect of the global state.

3. Discussion

In this section, we compare our approach to aspects of related work that haven't been considered through out the paper.

The Affective Reasoning Project uses an appraisal mechanism with two main steps [Elliott 1993][Elliott and Siegle 1993][Elliott 1994]: first, a set of appraisal frames is used to create the agent's interpretation of the situation, this interpretation is represented as a set of Emotion Eliciting Condition Relations (EECR); second, the left hand side of emotion eliciting productions is matched against the EECRs. Whenever a match is found the corresponding production fires, generating an emotion. This process has two important characteristics. First, more than one emotion may be generated; second, the emotion generated depends on the agent's interpretation of the situation, not on the actual situation. This later property is thus responsible for context-dependent emotions. In our approach, the first step is replaced by the automatic (as opposed to deliberative) effects of the context-dependent long term memory underlying our architecture. The emotion response depends on the cognitive representations currently more accessible in long term memory. This context-dependent emotion mechanism doesn't involve the symbolic processing required by the construal step in the approach of the Affective Reasoning Project, it just depends on the current pattern of activation of nodes in long term memory. As suggested in section 2, we may have several distinct appraisal processes (in sequence or in parallel), therefore our proposal may also generate several emotion-signals in response to the same initial situation. The main point to be retained is the view of emotion as an iterative process that explains different manifestations of an emotion, occurring one after the other.

Another parallel can be drawn between our model and the Affective Reasoning model.

Using our model as a reference, the construal step in the Affective Reasoning (i.e., using appraisal frames to build the EECRs) would be a preliminary cognitive appraisal stage whose results (the EECRs) would be stored in episodic memory; whereas the use of appraisal rules on the EECRs would be an affective appraisal stage following the earlier cognitive appraisal stage. This parallel would mean the Affective Reasoning model would capture those emotion processes that start with a cognitive appraisal stage.

Dolores Cañamero [Cañamero 1997] presents a model of emotion concerned only with motivation intensification. Like ours, her model is an architecture-level model that simulates the biochemical processes involved in emotion. Although this is a really appealing model there is no a priori reason to believe that computer agents should implement emotion in a biochemical-like fashion. We chose an information-processing based psychological approach instead because it deals with computer-like concepts. Besides, our model accommodates a broader variety of emotional phenomena than motivation intensification.

Some authors [Elliott 1993][Elliott 1994][Numao et al 1997][Reilly 1997] put the emphasis of their work at the task level. They try to use emotions to improve the performance of the agent in specific classes of tasks, such as natural language understanding, music arrangement, intelligent tutoring systems and interactive drama. This approach while certainly more focused on short and medium term development goals, may fail to create general intelligent agents with more flexible architectures, capable of better performance in any kind of task. In this respect we feel closer to Aaron Sloman and the Cognition and Affect Project [Sloman 1987][Sloman 1995]. However, while the Cognition and Affect Project has tried to imitate human emotions and personality, our research goal is different: we try to build mechanisms that can play the same roles in artificial agents as those responsible for the success of human beings. Considering the differences in supporting technologies, there is no reason to believe these mechanisms should mimic their natural counterparts.

Results

This paper proposes a definition of emotion, a classification scheme based on the proposed definition and a concrete implementation of emotion according to the definition. Definitions are not subject to experimentation therefore one cannot expect experimental results in the usual sense of the term. We have already shown in section 2 that the definition proposed can be used as the basis for an emotion classification scheme that unifies other classification proposals. In the last subsection of section 2, we have shown that the Salt & Pepper architecture for autonomous agents can easily implement the theory of emotion presented here. [Botelho and Coelho 1998b] shows that the framework just described lends itself to the clear definition of three classes of emotion learning: learning or updating appraisal structures, learning new emotion-signals and learning or updating emotion-responses. In this subsection we show that the proposed definition of emotion has the capability of explaining several interesting findings usually related to emotion. We have yet to show if an agent that exhibits emotion processes that comply with our definition performs better than another agent but this is not the goal of the present work.

Emotion is a process that starts with an appraisal phase and may include an emotion response. We believe that this might be an iterative process in which the earlier appraisal step gives rise to the initial component of the emotion. Then, the earlier emotion response changes the global state of the agent (externally and/or internally). It is possible that the appraisal of the new global state adds new components to the whole emotional process. Any of the appraisal stages involved in emotional processes may be either affective or cognitive.

This stepwise view of emotions explains several issues regarding emotions. First, it explains why some emotions as grief constantly interrupt one's current thinking [Sloman 1995]. According to the proposed model the repeated interruption of someone's thinking would be produced if several attention-shift warnings were generated in the appraisal stage. If some of this warnings give rise to the selection of a certain node to working memory, the current thinking would be interrupted. Several attention-shift warnings would be produced due to the appraisal of several external stimuli and internal memories that can occur in situations such as grief.

The proposed model also explains why some emotions give rise to other emotions. For instance, one may feel self-disappointed upon realizing that one has been dominated by anger. We can also understand why, in some cases, we quickly react to situations without thinking and, only after this first reaction stage, do we start feeling the emotional content that, in turn, may lead to another response stage. This might be the case when "all the manifestations of fear arise many seconds after slamming on one's brakes to avoid an auto accident"[b]. First, there is an affective appraisal of the situation triggering the reflexive behavioral response ("slamming on one's brakes"). At an information processing level of analysis, all the manifestations of fear arise later because affective appraisal is not conscious and reflexive actions don't involve deliberative processes. After the initial stage, one feels fear because the new global state is cognitively appraised -- this involves a conscious process. This later conscious stage has a pedagogical adaptive purpose, if we may say so. One has the opportunity to learn from the experience. Emotion responses to the late cognitive appraisal might include pulling oneself together, building a rationalization of the event, and learning how to avoid incurring in future similar situations.

Although this is only a conjecture, we feel that higher order emotions as euphoria and anxiety are complex processes involving the intervention of multiple appraisal stages and emotion responses. In this spirit we think the whole emotional process involved in the "auto accident episode", from the first appraisal and emotion response ("brake slamming") until the conscious experience of "all manifestations of fear" should be called fear (not only the last cognitive part of the process). This view is more consonant with the wide spread idea that many of our fast reactions are mediated by emotion.

Conclusions and future research

We have presented a multi-dimensional matrix classification scheme for emotions that integrates several other classification schemes proposed in the literature about emotions. We have emphasized the roles of and the processes involved in emotion in an architecture for autonomous agents. This is a radical departure from the mainstream view that addresses emotions at the task level (e.g., the Affective Reasoning Project and the Oz Project) and also from other lines of research that aim at creating artificial human-like characteristics such as emotion and personality (e.g., the Cognition and Affect Project).

We have also introduced a preliminary theory regarding the emotion generation process, from appraisal to behavior. This has focused on the appraisal process, mainly on the distinction between affective and cognitive appraisal.

We have illustrated our ideas about artificial emotions in the Salt & Pepper framework. We have discussed and implemented the affective appraisal stage and the behavioral stage of emotion. Special attention was paid to emotions in the intersection of attention shift warnings and performance evaluators, on one hand, with reflexive actions, motivator generators and plan selectors, on the other hand.

We have presented a partial implementation in a simple example of the application of artificial emotions in a concrete domain.

Although we haven't proposed anything new with respect to the emotion eliciting conditions, we have in fact stressed that we'd be better off if the motives of the agent were only implicitly used in the appraisal process. Reasoning chains involving lots of rules should be compiled into smaller more specific productions. Evaluations of the state of satisfiability involving explicitly represented motives should only be used when no specific rules are available. Using rule compilation techniques in the affective appraisal process, learning new emotion-signals and learning more and better emotion responses was discussed in detail in [Botelho and Coelho 1998b].

Of course much has yet to be done regarding artificial emotions. What does it mean for an artificial agent to feel an emotion? Is it useful to implement several degrees of emotion awareness in agent architectures? What kinds of emotions should be built in artificial agents? What artificial mechanisms can mirror the cognitive adjustments that are thought to take place when someone is feeling sorrow for a lost one? How can we build artificial agents with emotional empathy?

4. References

[Arafa, Charlton and Mamdani 1998] Arafa, Y.; Charlton, P.; and Mamdani, A. (1998) "Modelling personal service assistants with personality: from metaphor to implementation", *Proceedings of the Workshop Grounding Emotions in Adaptive Systems, International Conference on Simulation of Adaptive Systems (SAB98)*

[Aubé 1998] Aubé, M. (1998) "Designing adaptive cooperating animats will require designing emotions: expanding upon Toda's urge theory", *Proceedings of the Workshop Grounding Emotions in Adaptive Systems, International Conference on Simulation of Adaptive Systems (SAB98)*

[Balkenius and Morén 1998] Balkenius, C. and Morén, J. (1998) "A computational model of emotional conditioning in the brain", *Proceedings of the Workshop Grounding Emotions in Adaptive Systems, International Conference on Simulation of Adaptive Systems (SAB98)*

[Bates et al 1992] Bates, J., Loyall, A.B. and Reilly, W.S. (1992) "An architecture for action, emotion and social behavior", *Proceedings of the 4th European Workshop on Modeling Autonomous Agents in a Multi-Agent World (MAAMAW'92)*

[Beaudoin and Sloman 1993] Beaudoin, L.P. and Sloman, A. (1993) "A study of motive processing and attention" in Sloman, A., Hogg, D., Humphreys, G., Partridge, D. e Ramsey, A. (eds.) *"Prospects for Artificial Intelligence"*, p229-238, IOS Press, Amsterdam

[Bechara et al 1997] Bechara, A.; Damásio, H.; Tranel, D.; and Damásio, A.R. (1997) "Deciding advantageously before knowing the advantageous strategy", *Science, 275:1293-1295*

[Botelho 1997] Botelho, L.M. (1997) "Realização de agentes inteligentes: uma abordagem inspirada nos modelos cognitivos de tomada de decisão" ["Building intelligent agents: an approach based on the cognitive models of decision making"], Ph.D. dissertation, Department of Information Sciences and Technologies, ISCTE

[Botelho and Coelho 1995] Botelho, L.M. e Coelho, H. (1995) "A schema-associative model of memory", *Golden West International Conference on Intelligent Systems (GWICS' 95)*, p81-85

[Botelho and Coelho 1996a] Botelho, L.M. and Coelho, H. (1996) "Information processing, motivation and decision making", in Ein-Dor, P. (ed.) "Artificial Intelligence in Economics and Management" (AIEM' 96), p233-250, Kluwer Academic Publishers, London

[Botelho and Coelho 1996b] Botelho, L.M. and Coelho, H. (1996) "Emotion-based attention shift in autonomous agents", in Müller, J.P., Wooldridge, M.J. and Jennings, N.R. (eds.) *"Intelligent agents III. Agent theories, architectures and languages" (ATAL96)*, p277-291, Springer-Verlag

[Botelho and Coelho 1996c] Botelho, L.M. and Coelho, H. (1996) "Agents that rationalize their decisions", *ICMAS'96*, p3-10

[Botelho and Coelho 1998a] Botelho, L.M. and Coelho, H. (1998) "Artificial autonomous agents with artificial emotions", *Proceedings of the Second International Conference on Autonomous Agents (Agents'98)*, p449-450.

[Botelho and Coelho 1998b] Botelho, L.M. and Coelho, H. (1998) "Adaptive agents: emotion learning", *Proceedings of the Workshop Grounding Emotions in Adaptive Systems, International Conference on Simulation of Adaptive Systems (SAB98)*

[Bower 1981] Bower, G.H. (1981) "Mood and memory", *American Psychologist*, 36:129-148

[Cañamero 1997] Cañamero, D. (1997) "Modeling motivations and emotions as a basis for intelligent behavior", *Proceedings of the First International Conference on Autonomous Agents (Agents'98)*

[Damásio 1994] Damásio, A.R. (1994) "Descartes' Error: Emotion, Reason and Human Brain", Picador, London

[Elliott 1993] Elliott, C. (1993) "Using the affective reasoner to support social simulation", *IJCAI'93*, p194-200

[Elliott 1994] Elliott, C. (1994) "Research problems in the use of a shallow artificial intelligence model of personality and emotion", *AAAI'94*, p9-15

[Elliott 1998] Elliott, C. (1998) "*Story-Morphing* in the affective reasoning paradigm: generating stories semi-automatically for use with "Emotionally Intelligent" multimedia agents", *Proceedings of the Second International Conference on Autonomous Agents (Agents'98)*, p181-188

*[Elliott and Siegel 1993] Elliott, C. and Siegle, G. (1993) "Variables influencing the intensity of simulated affective states", AAAI Spring Symposium on Mental States.

[Forgas and Bower 1987] Forgas, J.P. e Bower, G.H. (1987) "Mood effects on person-perception judgements", *Journal of Personality and Social Psychology*, 53:53-60

[Goleman 1995] Goleman, D. (1995) “Emotional Intelligence”, Bantam Books, N.Y.

[Laird et al 1986] Laird, J., Rosenbloom, P. and Newell, A. (1986) "Chunking in SOAR: the anatomy of a general learning mechanism", *Machine Learning,* 1(1)

[Mayer et al 1992] Mayer, J.D., Gaschke, Y.N., Braverman, D.L. e Evans, T.W. (1992) "Mood congruent recall is a general effect", *Journal of Personality and Social Psychology,* 63:119-132

[Numao et al 1997] Numao, M., Kobayashi, M. and Sakaniawa, K. (1997) “Acquisition of human feelings in music arrangement”, *IJCAI’97*, p268-273

[Ortony et al 1989] Ortony, A., Clore, G.L. and Collins, A. (1988) “The cognitive structure of emotions”, Cambridge University Press, N.Y.

[Reilly 1997] Reilly, W.S. (1997) “A methodology for building believable social agents”, Agents’97, Marina del Rey, CA, USA

[Ritter 1993] Ritter, F.E. (1993) “Three types of emotions that can occur in a cognitive architecture like Soar”, *Workshop on Architectures Underlying Emotion,* University of Birmingham

[Simon 1967] Simon, H.A. (1967) "Motivational and emotional controls of cognition", *Psychological Review,* 74:29-39

[Sloman 1987] Sloman, A. (1987) “Motives, mechanisms and emotions”, *Cognition and Emotion,* 1:217-234

[Sloman 1991] Sloman, A. (1991) “Why consciousness is not worth talking about (yet)”, Proceedings of the International Conference on Cognitive Science (ICCS’91)

[Sloman 1995] Sloman, A. (1995) “What sort of control system is able to have a personality?”, in Trappl, R. (ed.) *Proceedings of the Workshop on Designing Personalities for Synthetic Actors*

[Srull and Wyer 1986] Srull, T.K. and Wyer, R.S. (1986) "The role of chronic and temporary goals in social information processing", in Sorrentino, R. and Higgins, T. (eds.) "Handbook of Motivation and Cognition: Foundations of Social Behavior", 1:503-549, Guilford Press, N. Y.

[Vilhjálmsson and Cassell 1998] Vilhjálmsson, H.H. and Cassell, J. (1998) "BodyChat: autonomous communicative behaviors in avatars", *Proceedings of the Second International Conference on Autonomous Agents (Agents'98),* p269-276.

---

[a] In this sense, "know" doesn't necessarily mean to be conscious, it means that certain information has to be implicitly or explicitly available to some component of the information processing mechanism of the agent.

[b] Quoted from [Elliott 1994] in which Clark Elliott refers to a question posed by Jeffrey Gray at the Workshop on Architectures Underlying Motivation and Emotion, Birmingham, England, 1993