Technical Report of
Departamento de Ciências e Tecnologias da Informação do ISCTE (1996)
Av. das Forças Armadas, Edifício ISCTE
1600 Lisboa, Portugal

An architecture for autonomous agency

Luis Miguel Botelho
Computer Science Department of ISCTE
Lisbon, Portugal
luis.botelho@iscte.pt

Helder Coelho
Helder.Coelho@di.fc.ul.pt
Department of Computing Sciences
Faculty of Sciences of the University of Lisbon
Lisbon, Portugal

Abstract

We present a general architecture for autonomous artificial agents that allows them to cope with the occurrence of internal and external stimuli simultaneously to their ongoing cognitive activities. This new architecture develops general proposals by Simon [1967] and Sloman [1987; 1995] particularized to the model described in [Botelho and Coelho 1995]. We relate the dispute between the primacy of affect and the primacy of cognition to the proposed architecture. Within the mentioned architecture we describe a new mechanism based on emotion and attribution, for controlling attention shift. Automatic as opposed to thoughtful and deliberate decision is a fundamental asset of our approach. Such automaticity is only possible because it draws on properties of the SALT model of memory [Botelho and Coelho 1995]. We relate our proposal to previous work on attention shift [Beaudoin and Sloman 1993], dynamic context-dependent change [Maes 1989] and commitment policies [Pollack and Ringuette 1990] and [Kinny and Georgeff 1991], and show how these former mechanisms may be defined within the proposed architecture.

1 - Introduction

Attention shift refers to what happens when some agent stops attending to its current thinking and attends to some external or internal event or object.

Consider a first situation (Example 1.1) in which a man fully absorbed in his thoughts trying to decide what gift to buy for his wife's birthday, while inattentively crossing a street with little traffic. Imagine this man suddenly senses something that makes him jump backwards avoiding being knocked out. "Oh, it's only a bicycle!", he thinks with surprise.

Consider now a second situation (Example 1.2) with the same man crossing the same street absorbed in the same thoughts while an equally absorbed friend passes him unnoticed, without interrupting his thoughts.

It's only natural any of the situations in examples 1.1 and 1.2 may has happened or is still to happen to the reader. It's also natural the reader may ask the same questions we have asked to ourselves. Why has the man interrupted his thoughts in Example 1.1, but not in Example 1.2? What mechanism mediates interruption in the first situation? Those and related questions constitute the topic of this paper.

Both stories in examples 1.1 and 1.2 are instances of the problem of attention shift. In both examples a man is paying attention to his current thoughts. In Example 1.1 the man interrupts them and shifts his attention to something else that made him jump (at this point we refrain ourselves from saying what exactly has he paid attention to). After having shifted attention, the man somehow reacts promptly. We won't be much concerned with his reaction, and only a suggestion is made in section 4.2.1 (Procedure 4.1). In Example 1.2, the man doesn't shift his attention and proceeds his current thinking.

Attention shift is a fundamental issue when we are considering the construction of autonomous artificial agents with limited rationality evolving in dynamic environments (e.g., [Agre and Chapman 1987], [Brooks 1991a], [Brooks 1991b], [Kirsh 1991], [Clancey 1993]). One aspect of limited rationality is the fact that reasoning takes time, another aspect is that both artificial and natural agents possess a limited amount of cognitive resources. Attention is such a limited resource. If the man of examples 1.1 and 1.2 was not limited in his rationality, attention shift wouldn't be such an important issue. First, the man in Example 1.1 would be able of recognizing a bicycle was moving towards him without disturbing his thinking; the man in Example 1.2 would have noticed his friend was passing by and would probably have talked to him, also without disturbing his thinking. Second, his decision regarding what gift to buy would have been made instantaneously.

A considerable amount of work reported in the Artificial Intelligence literature is related, in one way or another, with the problem of attention shift although in different guises. That work may be organized in two main categories: general architectures and specific attention shift mechanisms.

The first group addresses general agent architectures and design principles suitable for coping with the problem of attention shift under limited rationality (e.g., [Simon 1967], [Sloman 1987; 1995], [Bratman et al., 1988], [Georgeff and Ingrand 1989], [Pollack 1992]). In this respect, we present an agent architecture based on Simon' s [1967] and Sloman' s [1987] ideas, adapted to the options made in [Botelho and Coelho 1995]. An old dispute between Zajonc [Zajonc 1980], [Zajonc 1984] and Lazarus [Lazarus 1984] known as the primacy of affect vs the primacy of cognition dispute is discussed within our framework.

The second group describes specific mechanisms that may be used to control attention. In this latter group, some present filter-overriding and commitment policies (e.g., [Pollack and Ringuette 1990], [Kinny and Georgeff 1991], [Beaudoin and Sloman 1993]), and others describe cognitive mechanisms that lend themselves to cope with the attention shift problem (e.g., [Maes 1989], [Botelho and Coelho 1995]). In our view, although all the policies and mechanisms presented suit certain kinds of situations, none of them is capable of handling properly the situation described in Example 1.1. In this paper we present a mechanism of attention shift based on emotion and attribution. In concrete, we put forth the idea that when an agent experiences an emotion, it tries to determine what is the source of that emotion. If the agent attributes its experience to an external stimulus, then its current thinking is interrupted and attention is shifted to the external stimulus. If the agent attributes its emotion to its current thinking then no attention shift should occur. In order for this decision process to be fast enough, it relies on heuristic information and it is prone to errors. However, the same kind of errors has been observed in humans and has been termed the "fundamental attribution error" or the "correspondence bias" (e.g., [Ross 1977]).

The remaining of the paper is organized as follows. In section 2, we describe general design principles and an architecture for autonomous artificial agents with limited rationality. In section 3, we explain why the specific policies and cognitive mechanisms so far presented are not suited to handle all kinds of situations, and present some arguments in favor of our approach. We also show that those former attention shift mechanisms may be defined within our architecture. Section 4 describes the definition of attention shift mechanisms based on emotion and attribution within the general framework presented in section 2. Finally, section 5 presents some remarks and concludes showing how the proposed machinery can provide a basis for reinforcement learning.

2 - An architecture for autonomous artificial agents

Perhaps the most influential paper addressing the general problem of coping with multiple simultaneous needs is Herbert Simon' s "Motivational and emotional controls of cognition" [1967]. In his paper, Simon defends a serial cognitive mechanism whose current processing can be interrupted by the occurrence of external and internal events. About twenty years later, Aaron Sloman [1987; 1995] develops the ideas put forth by Simon in trying to specify a set of design principles for artificial agents. There are some differences between the two works. While Simon proposes an interruption mechanism based on some hierarchy of goals, Sloman proposes that motivators do not interrupt the current processing of the agent if they don' t penetrate a set of attention filters, even if they are very important, urgent or intense. Besides, Sloman explicitly refers the existence of motivator generators (as well as motivator comparators, generator generators, and so on), an idea omitted in Simon' s paper. However the main ideas of both works are the same, as they regard interruption mechanisms and emotion: emotion appears when the current cognitive program to achieve some of the agent' s goals is interrupted by another program to handle an external stimulus. The same kind of idea is also shared by others. For instance, Srull and Wyer [1986] say that affective states may arise of the interruption of current goals.

The architecture we present in this paper differs from Simon' s and Sloman' s ideas in two respects. First, our architecture relies on a parallel model of computation instead of a serial one (section 2.2). Second, besides the interruption-generated emotion, we argue that emotion may be a source of interruptions (section 4).

In this section we point some design principles for autonomous artificial agents drawn upon an analysis of examples 1.1 and 1.2. Those principles guide the specification of an architecture for autonomous agents based on previous work by Simon [1967] and Sloman [1987; 1995] adapted to the cognitive model of Botelho and Coelho [1995].

2.1 - Design principles

A general property involved in both situations of examples 1.1 and 1.2 is the ability to carry out parallel information processing activities. In both cases the man is simultaneously processing information related to his current thoughts and processing input information. Further more, the processing of input information doesn' t have the same status as the processing involved in thinking. Notice that the processing involved in thinking goes on consciously, while the processing of input information goes on unconsciously. This will become more clear if we remember that in example 1.2, the man didn' t noticed a friend that passed him by. The reader may argue that this was the case because the man wasn' t even processing input information, not because this is an

unconscious process. For the moment we will only mention two reasons against this argument. First, it is more simple to assume both processes go on in parallel and postulate the existence of a communication channel between them, than to assume the input information processing is turned on and off from time to time. Second, because there is a bulk of experimental psychological evidence that stimulus material previously presented to subjects below the level of conscious awareness, affects future performance on cognitive tasks by the same subjects, e.g. [Roediger 1990] and [Murphy and Zajonc 1993], lending some support to the hypothesis of the non conscious input information processing. We call a process that runs unconsciously a pre-attentive process, and a process that runs consciously an attentive or conscious process.

Any piece of information that is being processed pre-attentively, by a particular agent, at a certain point in time, may gain the agent' s attention, at another point in time, and become the object of an attentive process. That' s what happens in example 1.1, when the man senses something and jumps backwards -- he becomes conscious of that something, in spite he doesn' t know what exactly is happening. Since attention is a limited resource, the currently attended process must be interrupted in order to allow another process to gain attention. This means the agent' s architecture must provide interruption mechanisms. Besides interruption mechanisms the agent' s architecture must have interruption policies based on the stimulus being processed pre-attentively, but also on the state of the current attentive process, because interruption shouldn' t depend solely on the new stimulus.

In Example 1.1 the man sensed something, but only after jumping backwards did he realized it was a bicycle. It is worth noting that in spite he sensed something that made him jumped, the man was not conscious that a bicycle was moving towards him. The man must have heard a sound he was unable of recognizing as one produced by a bicycle. This sound must have caused him a particular emotion that made him jump. In order for the whole picture to make sense, we must assume certain postulates. First, some components of emotions may be produced faster than some components of cognitions, otherwise the man in Example 1.1 would have become consciously aware of the bicycle before he had experienced the emotion. Second, emotional states may be driven by some properties of the cognitive system, otherwise we can' t assume a causal relation between the sound the man heard and the experience of the emotion. Third, emotions may contribute to determine behavior, otherwise the man wouldn' t have jumped. Fourth, emotions have some sort of representation in the cognitive system, otherwise one wouldn' t feel the

emotion[1]. The previous analysis implies the interruption mechanisms discussed above should also depend on the relationship between the current attentive process and the experience of emotions produced either by actual events or by anticipation.

In example 1.1, the emotional state experienced could have been produced either by the man' s current thinking (he might have anticipated his wife generous reaction to his offering her the gift) or by the external stimulus. In face of this, the agent' s architecture must provide a fast attribution mechanism that identifies the source of the emotion.

2.1.1 - A word about affect and emotion

Along the text we'v e been using the word "emotion" in spite some may argue "affect" would have been a better choice. We have chosen to do so mainly for two reasons.

First, the property being referred by the word "emotion" is not easy to classify. On one hand, affect is generally used to refer to psychological states of feeling good or bad, that is, to psychological states with positive or negative valence; the states referred to by the word "affect" also don' t change quickly, instead they tend to last for relatively long periods of time; finally, it is difficult to identify the source of affect, that is, affect is not a localized state. On the other hand, emotion refers to psychological states much more specific than affect (e.g., fear, anger, joy, calm) all of which may also be characterized by a positive or negative valence (e.g., fear is negative, and joy is positive); contrary to what happens to affect, an emotion may come and go rather quickly; finally it is usually more easy to identify the source of an emotion than the source of an affect (i.e., emotions are localized states). The phenomena being described may change quickly. Actually, the man in example 1.1 suddenly sensed something and quickly jumped backwards. This pushes use more closely to emotion than to affect. Besides, it must be easy to identify the source of what is sensed. This too, counted as an argument in favor of term "emotion". Finally, each specific emotion leads to specific behavioral responses. That' s what happened when the man jumped backwards.

Second, at the eyes of the average reader, it would be more difficult to accept that the man in example 1.1 sensed a strong affect than that he sensed a strong emotion.

2.1.2 - Summary of design principles

In conclusion of the preceding analysis, we present the following design principles:

---

[1]The experience of an emotion is a cognition [Laird and Brestler 1991].

(1) The agent' s architecture must provide the ability to carry out parallel information processing activities. Some of those processing activities go on consciously (attentive processes), while others go on subconsciously (pre-attentive processes).
(2) Any piece of information that is being processed pre-attentively, at a certain point in time, may become the object of an attentive process, at another point in time.
(3) The agent' s architecture must provide interruption mechanisms, as well as interruption policies based on the stimulus being processed pre-attentively, on the waxing and waning of emotions, and on the state of the current attentive process.
(4) Some components of emotions may be produced faster then some components of cognitions.
(5) Some features of emotions may be driven by some properties of the cognitive system.
(6) Emotions have some sort of representation in the cognitive system.
(7) The agent' s architecture must provide a fast attribution mechanism that identifies the source of emotions.

The next step consists in defining an architecture for autonomous agents built according the preceding principles.

2.2 - More salt to SALT

In this section we extend the SALT model [Botelho and Coelho 1995] in order to define an architecture for autonomous artificial agents designed according to the principles outlined in section 2.1.

2.2.1 - A word about SALT and affect[2]

SALT is a model of memory for artificial agents originally described in [Botelho and Coelho 1995] and further developed in [Botelho and Coelho 1996a; 1996b; 1996c; 1996d]. SALT describes Long Term Memory as an associative network whose nodes contain declarative and procedural symbolic structures. Nodes are further characterized by an activation value which represents its accessibility in Long Term Memory. The greater the activation of a node the more accessible it is in memory. Each time Long Term Memory is searched by the agent to handle a given situation, nodes more activated are sampled first. The first node whose contents match the features of the situation is

[2]Here, the word "affect" is taken in a broader sense that includes the notions of affect and emotion, as described in section 2.1.1.

selected. An arc from node N_1 to node N_2 represents an association between the two, and the label of the arc represents the strength of that association. When a node is selected it receives a fixed amount of activation per time period. The activation received by any node spreads to the rest of the network through the arcs getting out of it, changing the accessibility of other nodes. This model exhibits interesting features including context-dependent behavior, and certain adaptive qualities.

For the current matter, the important feature to focus is the way an association is formed between two nodes. According to SALT the strength of the association from node N_1 to node N_2 depends on the relative frequency of selection of N_2 immediately after N_1 has been selected. When the strength of an association is greater than 0 we say an association has been formed. As an example, let us consider how a given node (e.g., a representation of a loved one) may become associated with positive affect. In order for an affect to become associated with any other node, it has to be represented by a node in Long Term Memory, thus the first problem is how do affects get represented? As with any other stimulus, SALT doesn' t make any assumptions regarding the formation of their representations in memory. However, since primitive emotions and affects exist in a very restricted number (e.g., [Ortony et al. 1988], [O' Rorke and Ortony 1994]), we may suppose people are born with their innate representations. More complex learned emotions would be represented through the combination of primitive emotions and affects. Now, we may turn to our initial problem: how is an association formed between a given node, N_1, and a node representing an affect, N_2? Since we have postulated the existence of nodes in memory representing affects and emotions, we just need to assume those nodes get activated whenever an affect or an emotion is produced. Hence, the strength of an association from a given node, N_1, to a particular affect or emotion, N_2, depends on the relative frequency with which the emotion or affect represented by N_2 follows the selection or occurrence of N_1.

2.2.2 - Structural Components

The extended SALT model (figure 2.1) relies on three main components: a cognition engine, an affect engine[2] and an Interruption Manager.

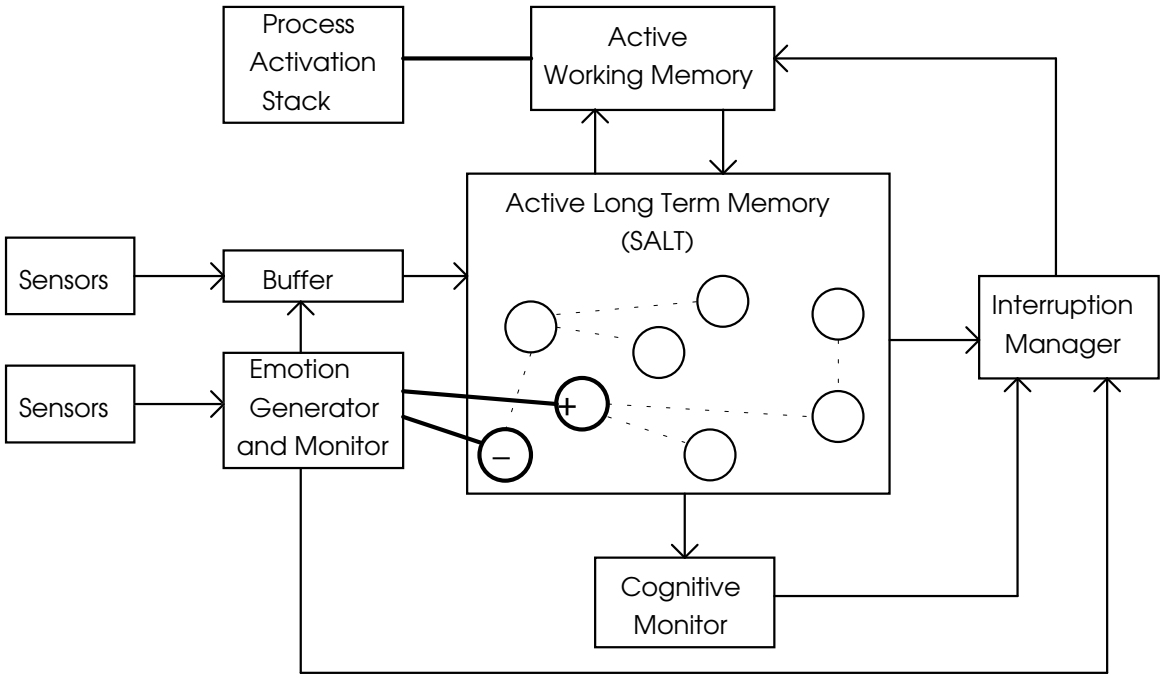

Figure 2.1 - Autonomous Agent Architecture

The cognition engine includes the following sub-components: an active Long Term Memory where all information is permanently stored; an active Working Memory where selected representation structures are manipulated; a set of Sensors that sample the outside environment, producing stimulus material; a Temporary Buffer where internal and external stimuli are placed; a Cognitive Monitor that produces a signal whenever a Long Term Memory representation becomes more activated than the representation being attentively processed in Working Memory; and a Process Activation Stack that keeps the status of interrupted processes.

The affect engine consists of a set of sensors that sample the outside environment (and physiological processes, in case of embodied agents); and an emotion generator and monitor that produces emotions as a result of the evaluation of external and internal information, directly activates affects and primitive emotions, and places representations of the emotions produced in the Temporary Buffer (as if they were external stimuli[3]). Finally, the Interruption Manager receives signals from the monitors specifying the node that should be considered to gain the agent' s attention, decides whether or not to interrupt

[3]Pedro Ramos suggested us this way of operationalizing the distinction and coupling of cognition and affect. However, shall this view reveal erroneous, the authors assume full responsability.

the agent' s current thinking using the most activated attention shift policy, and signals Working Memory accordingly.

Working Memory is considered an active component because, besides being used as a temporary store of information, it also processes the information stored. In the architecture presented, Working Memory is the locus of consciousness, that is, any piece of information copied from Long Term Memory to Working Memory, or produced as the result of a symbolic processing in Working Memory becomes conscious for the agent.

Long Term Memory is also an active component for it performs some processes, including the activation of nodes that match stimulus information stored in the Temporary Buffer, updating the strengths of associations between nodes that match stimulus information or between nodes that are selected to Working Memory, and updating the accessibilities of nodes according to their activation. The nodes in Long Term Memory are also active components since, whenever they receive activation, they send it to the nodes to which they are associated.

The Temporary Buffer is a queue with capacity for a fixed quantity of stimuli. When the capacity of the Temporary Buffer is exceeded, new stimuli override previous buffered stimuli. In the course of this process, it is possible that some stimuli are not processed. By default the Temporary Buffer obeys a first in first out discipline. However, the emotion generator may place internal stimuli in the front of the queue, to enable them to be processed first.

The Process Activation Stack is a "last in first out" data structure with limited capacity for a fixed quantity of process activation records. This ensures the first processes to be lost in case the stack exceeds its capacity, are the older processes; this also ensures that, while attention doesn' t shift, when the processing of a particular node in Working Memory halts, the first process to be resumed is the one whose activation record is on the top of the stack.

The sensors of the affect engine extract only a very restricted set of features suited to produce only a very restricted set of signals, such has "fast moving object", "loud noise" and "high temperature". The affect engine also monitors changes in the activation of the cognitive representations of affects and primitive emotions. As will become clear in section 4.2.2, this is extremely useful if we want emotions resulting from anticipation to interrupt the current attentive process. This may be done very rapidly, since the set of primitive emotions and affects is very small. If a sudden large variation of the activation of any of these representations is attributed to stimulus information being received, the evaluative component of an externally-driven emotion of expectation is produced, and the cognitive structure representing that emotion is directly activated. If the activation of this

emotion becomes the greatest of long term memory, a signal is sent to the Interruption Manager specifying this representation should be considered to gain the agent's attention.

The generation of an emotion is a gradual and distributed process. Earlier evaluations are produced very quickly by the Emotion Generator or anticipated by the Emotion Monitor. Signals are sent to the Interruption Manager. The Interruption Manager may interrupt the ongoing attentive process, which is another aspect of the emotion. The very interruption of an attentive process may produce another emotion. If the "whys" and "hows" of the emotion are inferred (e.g., who fired the gun, why has a plan failed, who made this goal achievable), then a more complex emotion description is produced. Complex emotion descriptions are placed in the Temporary Buffer, so that they may be processed by Long Term Memory. When a primitive emotion description is produced, the Emotion Generator automatically activates the node representing it. In both cases, the produced emotion may gain the agent' s attention. Whenever a description of an emotion gains the agent's attention it controls the agent's (overt and covert) behavior.

2.2.3 - General functioning of the model

Active Working Memory, Active Long Term Memory, the set of Sensors, the Emotion Generator and Monitor, the Cognitive Monitor, and the Interruption Manager all work in parallel, and produce asynchronous signals. The nodes in long-term memory send activation to each other, in parallel, too.

While the agent is processing its current thoughts in Working Memory, the sensors are collecting information from the environment (and also internal information, in the case of embodied agents, e.g., increased blood pressure) and placing stimulus information in the Temporary Buffer. At the same time, information in the Temporary Buffer is matched against nodes in Long Term Memory following a decreasing activation order, nodes are activated depending on the goodness of the match, and activation spreads to the network through the associations emanating from the nodes being activated. Simultaneously, the Emotion Generator receives information from its sensors, generates emotion descriptors (e.g., evaluative descriptions of the information with respect to the agent' s needs, motives, values, desires and goals - p.c.s.s and s.c.s.s, according to [Wright 1995] - along with the emotion sensor information) and sends control signals to the agent' s body (in case of embodied agents). If the emotion descriptor produced matches any of the affects or primitive emotions of the agent, the Emotion Generator directly activates the corresponding node in Long Term Memory. Notice that this process of direct activation is very fast since the Emotion Generator only activates a very restricted

set of nodes. Otherwise it places the evaluation description in the Temporary Buffer and it will be processed as a regular external stimulus.

The activation of nodes in memory is observed (in parallel) by the Emotion and the Cognitive Monitors. If any node in Long Term Memory becomes more activated than the node driving the current thinking of the agent, the Cognitive Monitor sends a signal to the Interruption Manager.

If the Emotion Monitor detects a sudden and large variation of the pattern of activation of the nodes representing primitive emotions or affects, and attributes this variation to the stimulus being processed, it generates an externally-driven emotion of expectation and activates the node representing it. If this node (or any other node representing primitive emotions or affects) becomes the most activated in long term memory, the Emotion Monitor sends a signal to the Interruption Manager, telling it this node is the most activated, at the moment.

When the Interruption Manager receives a signal specifying what node should be considered to gain the agent' s attention, it selects the most activated attention shift policy from Long Term Memory and decides whether or not the node gains attention. If the Interruption Manager decides to interrupt the current thinking of the agent, it tells Working Memory what node should be processed.

When Working Memory is interrupted and receives a signal specifying what node should be processed, it places the current process activation record on the top of the stack of process activation records, copies the specified node to Working Memory, and initiates its processing, or resumes it if the node copied corresponds to an activation record stored in the stack of process activation records. Active Long Term Memory ensures that the strength of the association from the node driving the current thinking and the node to be processed is updated according to their relative frequency of recruitment to Working Memory, as was the case in the original SALT model [Botelho and Coelho 1995].

When Working Memory is not interrupted and its current processing comes to an halt, the activation record on top of the activation stack is copied and the corresponding process is resumed. If the activation stack is empty, the most activated motivator (e.g., needs, goals, decision problems) of the agent is recruited from long term memory and processed.

The general functioning of the model will become more clear in section 4 in which an example is described.

2.3 - The primacy of affect vs the primacy of cognition

The dispute between Zajonc and Lazarus [Zajonc 1980], [Zajonc 1984], [Lazarus 1984] regarding which phenomena has primacy in human behavior is twofold: first, Zajonc argues that affective phenomena occur earlier than cognitive phenomena, while Lazarus argues in the opposite direction; second, Zajonc argues that affect and cognition are produced by separate systems.

What phenomena occurs first?

According to the architecture proposed in section 2.2, affective and cognitive processes take place in parallel. Both of these processes produce phenomena that influences the agent' s behavior. Sometimes a particular affective phenomenon occurs first, other times a particular cognitive phenomenon occurs first. However, the fact that an affective phenomenon occurred earlier in a certain chain of events does not preclude the possibility that a cognitive phenomena may occur before another affective phenomena, in the same chain of events. And the same may be said, starting with a cognitive phenomenon, *mutatis mutandis.*

Lazarus also complained that Zajonc' s stance arouse of a definitional problem, not of an objective analysis of human functioning. Of course, Zajonc maintained that Lazarus was the one with definitional problems. We think that it is very difficult to argue about definitional problems without grounding the argumentation in an accepted architecture. To give an example, one of the arguments of Lazarus was that any affective phenomena is always preceded by an encoding phase, that is, a cognitive task by definition. It is worth noting the architecture proposed in section 2.2 assumes the affect engine is equipped with its own set of sensors. If both Lazarus and Zajonc had agreed upon a particular architecture, this question would have been settled.

Are affect and cognition produced by separate systems?

Once more, the issue is settled, if we agree upon definitions of affect and cognition grounded in specific architectures. If we take the architecture proposed in section 2.2, we will find two systems: the so called affect and cognition engines. However, we will also notice those systems are not completely separable. In fact, both systems inform and are informed by Long Term Memory. As an example, consider that the affect engine may directly activate nodes in Long Term Memory but its functioning may be directed by the pattern of activation of particular nodes in Long Term Memory. It should also be noticed that the recruitment of nodes to Working Memory may depend on the activation of nodes by the affect engine. Another point of contact is the Interruption Manager: it receives signals from both the affect and the cognition engines, and both of them are informed by its operation. As an example, consider that attention shift is controlled by the Interruption

Manager, but if current thinking is related to the achievement of a certain goal, then a negative affective state may arise if the current thinking is interrupted.

Evidence supporting our hypothesis of the existence of different but not totally separate systems comes from neuroscience. Phineas Gage, the subject of the "Descartes' Error" by António Damásio [1994], suffered severe brain injuries that precluded certain forms of emotion-based behavior, while leaving untouched his reasoning capabilities. However, his decision making skills were affected by the lack of some emotion-based controls. That is, there are certain cognitive tasks (such as decision making) that are affected by emotions.

More evidence of the intertwining of affect and cognition comes from experiments on affect priming [Mayer et al. 1992] and [Forgas 1994].

3 - Major approaches to attention shift policies

Several proposals regarding mechanisms used to control the shift of attention have appeared in recent literature on Artificial Intelligence. Pollack and Ringuette [Pollack and Ringuette 1990], Kinny and Georgeff [Kinny and Georgeff 1991], and Beaudoin and Sloman [Beaudoin and Sloman 1993], among others, have presented commitment and filter-overriding policies explicitly directed to the problem of attention shift. Botelho and Coelho [1995], and specially Pattie Maes describe cognitive mechanisms that lend themselves to cope with the attention shift problem, although they don' t make explicit any policy directed to that problem. In the present section we consider each of these cases, in turn.

3.1 - Commitment

The approaches described in [Pollack and Ringuette 1990] and [Kinny and Georgeff 1991] are very similar and may be termed as commitment policies, although both works are based on previous general architectures (IRMA [Bratman et al., 1988] and PRS [Georgeff and Ingrand 1989]) which also used the expression "filter-overriding mechanisms".

[Pollack and Ringuette 1990] describes an agent that commits itself to its choices as a form of avoiding processing additional information, including external stimuli information. The agent moves around a rectangular grid with tiles (the Tileworld), obstacles and holes with the goal of filling the holes with tiles. Filling a hole yields a certain score, depending on the specific hole. The global performance of the agent is measured by the score it gets after a fixed period of time. Holes and obstacles appear and disappear randomly. The agent sets a plan in which it chooses to move itself to a certain

hole, along a certain path, avoiding obstacles and collecting tiles to fill the selected hole. The agent may only consider alternative plans while it is not filling a hole. This means the agent is committed not to pay attention to new stimuli (i.e., new holes and obstacles) in certain phases of its agency (while filling a hole).

evaluation

The described strategy has the main advantage of avoiding any evaluation of a stimulus in order to decide whether or not to interrupt the agent' s current processing, during the phases in which the agent is committed to its options. Therefore, this is a suited strategy in dynamic but benevolent environments. However, we think this strategy is not suited in challenging environments, because in such environments an endangering event may occur while the agent is committed not to pay attention to external stimuli.

The agents described in [Kinny and Georgeff 1991] evolve in the same world (with the simplification that there are no obstacles) and have the same goals as the agents of [Pollack and Ringuette 1990], but their commitment policy is somehow different. The agents of Kinny and Georgeff commit themselves not to attend to new stimuli while they have not performed a fixed quantity of planning steps.

evaluation

This commitment strategy is worse than the previous one since, the periods in which the agent doesn' t pay attention to external stimuli don' t even depend on the pre-judged value of the agency being performed during that period. Notice that the agents of Pollack and Ringuette don' t pay attention to external circumstances only when their activity has much value (as anticipated by the designer).

3.2 - Attention filter penetration

Beaudoin and Sloman [Beaudoin and Sloman 1993] present a mechanism of attention shift with some advantages over the commitment policies considered in section 3.1. In their theory (AFP, the Attention Filter Penetration theory) a stimulus is characterized by three parameters: insistence, importance and urgency. Insistence is the stimulus propensity to override the agent' s attention filter. The agent considers attending a new stimulus only if its insistence is greater than a certain threshold, that is, if the stimulus penetrates the attention filter. The stimuli that manage to penetrate the attention filter are further evaluated in terms of importance and urgency before gaining the agent' s attention. Stimuli that are not sufficiently important or urgent don' t gain the agent' s attention. The

AFP theory allows insistence, importance and urgency to be either fixed or the result of computations that may depend on current circumstances, in particular, on current beliefs of the agent.

evaluation

In abstract the AFP theory has some advantages and potential disadvantages when compared to the commitment policies discussed in 3.1, however its real value depends on the architecture within which it is set to work.

If new stimuli are not processed in parallel, they always interrupt the agent' s current thinking, even if it is resumed only a few instants latter. This is a potential disadvantage of the AFP theory.

If new stimuli are processed in parallel, the advantages and disadvantages of the AFP theory depend on the computations needed to calculate their insistence, importance and urgency.

If those parameters are fixed in advance then the decision of whether or not they should gain the agent' s attention is automatic and it takes a maximum time of three comparisons. However, from a conceptual point of view it seems a very unrealistic design option because both the insistence, the importance and the urgency of a stimulus depend (in general) of contextual circumstances.

If those parameters depend on the context they must be computed whenever they are needed. In this case the computation may take quite a while, but the result is more realistic. There is still a problem to be solved: there is no clear (or even possible) relation between the intuitive notion of insistence and the architecture proposed by Sloman [1987]. However, as will be seen in section 4, this relation is quite natural within the architecture described in section 2. In this framework, insistence is the activation of the representation of the stimulus in the agent' s Long Term Memory. It is worth noting that, with this interpretation, (i) insistence is computed automatically (as opposed to attentively, consciously, or thoughtfully), (ii) insistence has an adaptive nature because it is learned through the agent' s experience; (iii) importance and urgency are relative only to the most activated criteria (e.g., beliefs).

3.3 - Behavior activation mechanism

[Maes 1989] presents a mechanism of activation of behaviors based on the same set of ideas of the cognitive model described in [Botelho and Coelho 1995]. Both rely on the concepts of spreading activation, and association. However, the SALT model [Botelho

and Coelho 1995] was intended as a cognitive model used to support some forms of decision making in question-answering settings, not as a model of autonomous agency.

In the model proposed by Maes [1989], when the agent is faced with external stimuli, certain nodes in a network of behaviors are activated, and the most activated behavior whose preconditions are met, is performed. If new stimuli are presented to the agent, the pattern of activation of the network of behaviors changes accordingly, and another behavior is performed by the system.

evaluation

The mechanism proposed by Pattie Maes supports a restricted form of an insistence-based decision of attention shift. Although the behavior of the system depends on its dynamic adaptation to the changing environment, the system has no learning capabilities, since the strengths of the associations between behaviors in the network are fixed in design-time. This is a restriction to the concept of insistence of the AFP theory, when it is considered within the framework proposed in section 2. However, the automatic and dynamic nature of the system' s response to new stimuli should be emphasized.

The mechanism proposed by Maes presents another disadvantage when compared to the AFP theory, because the reaction of her system to a new stimulus does not depend on the stimulus importance nor urgency. In this way, it is impossible to block the interruption caused by insistent stimuli, even if they are not important or urgent.

In summary, we conclude the AFP theory is the up to date most comprehensive approach to the attention shift problem, if it is interpreted within the architecture described in section 2. Nevertheless it is not capable of explaining what happens in the situation described in Example 1.1.

4 - Emotion-based attention shift

In this section we show how the architecture proposed in section 2 provides all the means for implementing commitment as well as AFP policies of attention shift. We further present two new mechanisms of attention shift based on emotions. Finally we show how the architecture described in section 2 supports these two new mechanisms, resorting to the situation described in Example 1.1.

4.1 - Previous commitment and interruption policies

To see how commitment policies such as those of [Pollack and Ringuette 1990] and [Kinny and Georgeff 1991] as well as the interruption policies of the AFP theory [Beaudoin and Sloman 1993] are supported by our architecture (section 2), we will describe the conditions under which a new stimulus gains the agent' s attention, according to the proposed architecture.

When the Cognitive Monitor or the Emotion Monitor detects certain conditions (e.g., a node representing an emotion has become more active than the node driving the current thinking of the agent) a signal is produced telling the Interruption Manager what node should be considered to gain the agent' s attention. This is the equivalent concept of insistence and penetration, in the AFP theory.

When the Interruption Manager is informed that a certain node should be considered to gain the agent' s attention, it selects the most activated attention shift policy from Long Term Memory and uses it to decide whether or not the node gains the agent' s attention.

commitment policies

If we want an agent to use commitment policies the execution of a plan need only set an interrupt enable flag whenever an interrupt is allowed to occur. In the case of [Pollack and Ringuette 1990] this flag is set while the execution of the plan doesn' t reach a certain phase (e.g., filling a hole in the Tileworld). In the case of [Kinny and Georgeff 1991] this flag is set while the execution of the plan hasn' t performed a predetermined number of steps, yet. The agent must be equipped with an attention shift policy that interrupts the current processing whenever the Interruption Manager is informed of a new node to be processed and the interrupt enable flag is set.

It is worth noting the way commitment policies are defined as described above is not as pure as described in [Pollack and Ringuette 1990] and [Kinny and Georgeff 1991] because, in the present approach, a new stimulus interrupts the current attentive processing if the agent is not committed to ignore new stimuli and the node representing it has become the most activated in Long Term Memory. If we want any new stimulus to interrupt the agent' s current thinking whenever it is not committed to ignore new stimuli (disregarding its activation), then the Cognitive Monitor should signal the Interruption Manager whenever a new stimulus is detected.

By the above argument we conclude that the architecture proposed in section 2 subsumes the architectures proposed in [Bratman et al., 1988] and [Georgeff and Ingrand 1989], with respect to attention shift mechanisms.

The AFP theory

As described in [Beaudoin and Sloman 1993], the AFP theory relies on three properties of a stimulus: insistence, importance and urgency. We shall focus each of these in turn.

insistence

As we have already argued along the text and, in particular, in the beginning of this section (4.1), the architecture proposed in section 2 provides the mechanism of attention filter penetration based on the concept of insistence. In the present theory, insistence is grounded in certain conditions pertaining to the pattern of activation of Long Term Memory. We want to stress that the determination of insistence doesn' t involve any evaluation regarding the motives and believes of the agent. Instead it arises of a property of the automatic information processing of the agent. Further more, insistence is a global property dependent of the context, for the pattern of activation depends both on the current context, and on the history of past interactions of the agent, and the activation of a particular node is related to the activation of all nodes in Long Term Memory.

urgency

The urgency of a stimulus may be implemented in the present theory in a number of ways. First, the Temporary Buffer (section 2.2.2, figure 2.1) may be sorted by urgency. Second, very urgent situations (those related to the Emotion Generator) are handled very rapidly: primitive emotions and affects are directly activated by the Emotion Generator; some urgent conditions regarding the activation of primitive emotions and affects are immediately signaled to the Interruption Manager by the Emotion Monitor; and emotion descriptors that don' t match an affect or a primitive emotion are placed in front of the queue of the Temporary Buffer. Therefore, the intended effects of urgency are also provided by our architecture. Further more, some of these effects depend only on structural determinants, they don' t involve any computation to evaluate the urgency of stimuli: stimuli produced by the Emotion Generator become urgent by that very fact.

importance

The importance of a new stimulus may be represented by a fixed predetermined value, or by an expression to be evaluated when needed, in the node representing the stimulus. When the Interruption Manager is informed that a particular node should be considered to gain the agent' s attention, the importance-based attention policy compares the importance of that node to the importance of the node driving the current thinking of the agent and decides accordingly.

conclusion

The architecture proposed in this paper is compatible with the requirements of the AFP theory of attention shift. We further stress some of this requirements (insistence and

urgency) are handled by our architecture in automatic and sometimes structural (as opposed to deliberative and thoughtful) ways. That is, they are properties that emerge of the automatic information processing and of the architecture of the agent, not of some computation involving goals, beliefs and the like. In this respect, our architecture shares some of the ideas stressed by Rodney Brooks [1991a, 1991b], in particular his claim that (most times) one doesn' t need deliberative and thoughtful processes in order to get intelligence

We have already seen how the proposed architecture may handle any commitment or interruption strategies. Now we emphasize an agent may possess several of these possibly conflicting attention shift policies. Each of the attention shift policies may be represented within a different node. The activation of nodes in Long Term Memory depends on contextual and historic factors, that is, at any point in time there is only one such node that is the most activated. Since the Interruption Manager recruits the most activated attention shift policy at that time, only one policy will be used, exactly the one that is more appropriate in each set of conditions as learned by the agent along its previous interactions.

4.2 - Emotion-based attention shift

In section 4.1 we showed how the attention shift policies presented so far can be defined in the architecture presented in this paper. In this section, we propose two emotion-based attention shift mechanisms and relate them to the situation described in the introduction (section 1, Example 1.1).

4.2.1 - Attention shift by event-driven emotion

In the situation of Example 1.1 (section 1) the man reacted to environmental information without being aware of what was really going on. Only after having jumped backwards did he noticed a bicycle was moving towards him. The example only says that the man sensed something that made him jump avoiding being knock out. In this section we interpret the situation assuming the man felt an emotion and reacted has a result of that emotion. What may have happened in this situation, in terms of our architecture?

The man was thinking about a gift to offer his wife. We suppose this thinking was driven by the node "Chose Gift to Wife" with a certain activation, A_1, in Long Term Memory. At the time this was happening, the man' s cognitive sensors detect a bicycle was moving towards him, and place the encoding of this information in the Temporary Buffer. This information is matched against Long Term Memory and, when the activation of the network settles, the node "Moving Bicycle" has received the greatest amount of

activation due to the stimulus processing, say A_2. We now suppose A_1 is greater than A_2, therefore the node "Moving Bicycle" doesn' t interrupt the agent' s attentive processing. At the same time, the same stimulus information is being sensed by the emotion sensors of the agent. We may suppose the encoding produced by the emotion sensors can be read as "medium size moving object near agent". We also suppose that the Emotion Generator produces the evaluation "fear of moving object" which matches the node representing the primitive emotion "fear". If the node "fear" becomes the most active of the network, the Emotion Monitor signals the Interruption Manager that "fear" should be considered to gain the agent' s attention. Supposing the most activated attention shift policy decides this node gains attention, the Interruption Manager interrupts the active Working Memory, which in turn, places the current process activation record on top of the stack of activation records, copies the node "fear" and initiates its processing. Finally the processing driven by the node representing fear may be supposed to do something like the following (Procedure 4.1):

(1)	Activate an attention-shift policy that disables interrupts
(2)	Find out what is the cause of the emotion (Attribution) and act accordingly
(3)	If the cause of the emotion is not known, then perform the most activated default action
(4)	Activate a more permissive attention-shift policy

Procedure 4.1

The execution of the Procedure 4.1 may have two kinds of outcome. In the first case, we assume the cause of the emotion is present in the front of the queue of the Temporary Buffer, either because it is still there, or because the cognitive sensor has sampled the environment again and still finds the bicycle (Procedure 4.1, step 2). If the prescribed action is "jump backwards", then that' s what the man did. In the second case, we assume the cause of the emotion cannot be found. Therefore, the man performs the most activated default action. By that time, it is natural the most activated action for fear is to jump backwards, since the man is crossing the street which may have cause automatic activation of such a default action for fear (Procedure 4.1, step 3).

Why did the man become aware of the bicycle ("Oh, it' s only a bicycle!")? The explanation is that when the sensor sampled the environment again the bicycle was still there. Hence the node "Moving Bicycle" got activated again and, this time it gained the man' s attention. The man didn' t become aware of the bicycle before he jumped just because the prescribed action for the situation took place before the node "Moving

Bicycle" has gained the man' s attention. Notice that interrupts have been disabled while the man was reacting to fear. The final question is: if the node "Moving Bicycle" didn' t gain the man' s attention the first time, how come it gained it now? Well, in spite the activation decays with time, the node "Moving Bicycle" has received a certain amount of activation the first time the bicycle was sensed, and it receives an extra amount of activation when the man' s cognitive sensors found it again. The resulting activation was simply enough to make it the most activated node in the network.

4.2.2 - Attention shift by anticipation-driven emotion

In this section we propose another emotion-based attention mechanism. If (i) the pattern of activation of nodes representing primitive emotions and affects changes significantly in a short period of time, and (ii) the variation of activation is attributed to the stimulus being processed, then the agent anticipates an emotion or affect -- the one whose activation increased the most. When a positively valenced emotion or affect is anticipated and attributed to the external environment, a good externally-driven emotion of expectation is generated. When a negatively valenced emotion or affect is anticipated and attributed to the external environment, a bad externally-driven emotion of expectation is generated. If the node representing the externally-driven emotion of expectation becomes the most activated in long term memory, a signal is sent to the Interruption Manager. The rest of the process happens as in 4.2.1. In what follows we give a rationale for this proposition.

What does a large variation of the activation of a node representing an emotion mean? There are two possible reasons for such a large variation.

First, the node representing the emotion may have been directly activated, which means that an emotion was produced by the Emotion Generator, as the result of the evaluation of certain information. In this case, it seems reasonable that the node representing the emotion may be considered as a candidate for gaining the agent' s attention, since it is widely accepted that emotions play a special role in human functioning.

Second, other nodes may have been activated and may have sent a significant proportion of their activation to the node representing the emotion, which means the emotion may be anticipated (although, possibly not consciously). To see that this is so, we have to think about the meaning of an association from one node, N_1, to another node, N_2. According to the SALT model of memory [Botelho and Coelho 1995], if the association from N_1 to N_2 is very strong, it means that N_2 often follows N_1 (either because N_2 is recruited to Working Memory immediately after N_1, or because the

stimulus represented by N_2 appears after the stimulus represented by N_1). Therefore, a large variation of the activation of a node representing an emotion that results of the activation of a previous node may be regarded as an indication that the emotion will follow. If the anticipated emotion results from the ongoing conscious thinking of the agent, then it won' t be worth to interrupt it. However, if the anticipated emotion is attributed to an external event, it seems we have a good reason for considering interrupting the agent' s current thinking. Therefore, a large variation of the activation of a node representing an emotion may constitute a good reason to interrupt the agent' s current thinking, only if the cause of the variation is attributed to external information (maybe, there is danger or maybe a gorgeous girl is passing by). Hence, the agent needs a fast way of making such attributions.

The strength of the association from node N_1 to node N_2 is a good heuristic to assess the probability that N_1 has caused N_2, since the association will be stronger in case N_2 often follows N_1. Consequently, we propose to use the strength of associations as the basis for the attribution process. This has the great advantage of being an automatic process. In fact, the strength of an association is dynamically computed by the automatic information processing of the agent.

Let S-Env denote the strength of the association from the node representing the stimulus placed in front of the queue of the Temporary Buffer to the node representing the anticipated emotion; and S-Think denote the strength of the association from the node driving the current thinking of the agent to the node representing the emotion. If S-Env is greater then S-Think, then the cause of the anticipated emotion is attributed to the external environment.

The attention shift mechanism proposed in this section goes as follows: if the Emotion Monitor detects a large variation in the activation pattern of nodes representing affects and primitive emotions, and the cause of the emotion represented by the node whose activation increased the most is attributed to an external stimulus, then the Emotion Generator produces an externally-driven emotion of expectation, and the node representing the emotion is activated. If this node becomes the most activated in long term memory, it will be considered for gaining the agent' s attention (as in subsection 4.2.1). This mechanism considers relative variations of opposite emotions or affects. What counts is not the absolute variation of a certain emotion or affect, but its variation relative to the opposite emotion. As an example, what counts is the variation of the difference between the activation of the node representing happiness, and the activation of the node representing sadness.

Consider now how this new mechanism may also explain what happened in the situation described in Example 1.1. This time we may suppose that the man heard some sound he couldn' t identify with a bicycle. The node representing this sound ("Sound") becomes highly activated because it receives activation in two instants of time. First, it gets a certain amount of activation when the encoding produced by the cognitive sensor is placed in the Temporary Buffer and matched against Long Term Memory becoming the most activated node at that moment and gaining the agent' s attention (notice that we are assuming the man heard the sound before he jumped). Second, when it gains the agent' s attention it is copied to Working Memory and it is further activated. When the Interruption Manager tells Working Memory "Sound" should be processed, the activation record of its current thinking ("Chose Gift to Wife") is put on top of the Process Activation Stack, "Sound" is copied to Working Memory, and is processed. The agent becomes aware of the sound, but we assume nothing else happens (the sound was not identified with a moving bicycle, and no action is specified in the node "Sound"), therefore the processing of "Sound" ends, the activation record on top of the stack is popped down, and the thinking about choosing a gift proceeds. If we suppose the circumstances of the man walking across the street together with the large activation of the node representing the sound he heard produces a large variation of the activation of a negative valence emotion (e.g., fear) relative to the activation of the opposite positive valence emotion (e.g., boldness), then the Emotion Monitor may attribute this variation to the external stimulus (the heard sound), produce a bad externally-driven emotion of expectation and the node representing this emotion is activated. We suppose this node becomes the most activated in long term memory, thus the Emotion Monitor signals the Interruption Manager that node should be considered to gain the agent' s attention. The rest of the story is as hypothesized in the former explanation (section 4.2.1).

Having this mechanism what do we gain with the mechanism presented in section 4.2.1? The answer is twofold: first, the former mechanism relies on the operation of the Cognitive Monitor which cannot be avoided, second, this new mechanism cannot be applied to monitor variations of the activation of nodes representing complex emotions (i.e., not primitive emotions). The reason for this latter argument has to do with efficiency considerations. It is only conceivable to have a very fast monitoring device (the Emotion Monitor) if it takes care of only a few nodes -- the very restricted set of primitive emotions and affects.

5 - Concluding remarks

We have presented an architecture for autonomous agents with interruption mechanisms, within which several attention shift policies may be defined. We have proposed two new mechanisms of attention shift based on emotion (attention shift by event-driven emotion, and attention shift by anticipation-driven emotion). In both proposals, when certain conditions hold, it is possible that an emotion is felt by the agent (i.e., it gains the agent' s attention). It was also suggested that, besides having the capability of gaining the agent' s attention interrupting its current attentive processing, emotions further control the agent' s behavior. Procedure 4.1 (section 4.2.1) sketches the processing performed by the agent when a particular emotion gains its attention.

Attention shift by anticipation-driven emotion is an essential feature in building successful behavior directed at avoiding dangerous (or otherwise undesired) situations, since the agent is led to attend to its environment even before the dangerous event has really taken place.

Besides the new emotion-based interruption mechanisms described, we have showed how former interruption mechanisms and policies might be defined within the proposed architecture.

In this section we show that the proposed architecture provides the basis for reinforcement learning (section 5.1), and we present some final remarks in section 5.2.

5.1 - Reinforcement learning

Having shown the proposed architecture is capable of information evaluation regarding the agent' s motives, as well as attribution processes, it is only natural it provides enough basis for reinforcement learning. In fact, reinforcement learning results of attributing the experience of negative or positive affects to the outcome of the agent's actions.

Just to have a general flavor of the way reinforcement learning can easily be built within the proposed architecture, consider a situation in which an agent is faced with a decision problem (DP) and uses a long term memory node (N) to produce the decision outcome D. Suppose also that after having made the decision, the agent experiences a negative affective state attributed to D. Then, according to [Botelho and Coelho 1996c], the agent' s information processing mechanism (i) creates a new node (Avoid) containing a strong need for a desired outcome when facing decision problem DP, a goal of avoiding negative affect, and a belief that if decision D is produced then the goal of avoiding negative affect is not achieved; and (ii) associates the node N to the node Avoid. Therefore, in agreement with [Botelho and Coelho 1996a], when the agent is faced with another instance of decision problem DP, the node Avoid enables it to avoid producing decision outcome D, and hence experiencing a negative affect. That is, the agent learns to

avoid making decisions that produce negative affect, which is a form of learning by (internal) punishment.

5.2 - Final remarks

In the pervious section we have shown that our architecture provides a basis for reinforcement learning. In this section we point some other relevant features of the proposed architecture.

Although we consider that the fundamental information processing involved in the affect engine (sensoring, evaluation, monitoring, generation of representations) is of the same nature of that involved in the cognition engine, we have assumed a distinction based on the roles cognition and affect play on the agent behavior, on some architectural properties and on some differences between the specific information processing carried out by the two systems (e.g., the kind and quantity of features extracted by the emotion and the cognitive sensors).

The architecture proposed in this paper may exhibit all the requirements Herbert Simon [1967] postulated for intelligent agents. The two main points in Simon's proposal were a terminating mechanism for goals and an interruption mechanism.

It is clear that our architecture may be equipped with several mechanisms and policies of interruption.

According to Simon's ideas, goals should be terminated (i) when they become achieved; (ii) when they become achieved well enough (satisficed); (iii) when motivation or time is run out; and (iv) when they become believed to be impossible to achieve.

COMINT model [Botelho and Coelho 1996a] extends SALT to enable the termination of tasks under a continuum of motivation conditions. In this development, the agent's motivation to solve a given problem imposes termination conditions. If the motivation to search information relevant to the problem at hand is very strong, then all possible solutions to the problem are found. This is a termination condition equivalent to condition (i). When motivation to search information is not very strong, a satisficing termination criteria is met. However, if motivation to search is even weaker, the problem solving activity is terminated before the problem is actually solved. This corresponds to conditions (ii) and (iii). However yet another satisficing condition (condition ii) may also be produced by the COMINT model of decision making. That is, a problem solving task may be terminated when the solution found conforms to the agent's desires. Finally, as the motivation to search information may vanish whenever the agent finds the problem is impossible to be solved, the COMINT model of decision making also provides terminating condition (iv). In conclusion, if SALT is replaced by the COMINT model,

the architecture here proposed satisfies all requirements put forth by Herbert Simon in 1967.

It is also important to stress that most interrupt mechanisms and policies as well as terminating conditions are provided by automatic or even structural features of the proposed architecture, as opposed to deliberate and thoughtful processes. This asset meets Brooks' [1991a and 1991b] claim that intelligence can be achieved without (explicit) representations. In spite of this similarity with Brooks ideas, there are significant differences. First, Brooks' robots don' t really have the problem of attention shift as it has been described along this paper. In fact it is meaningless to establish a distinction between attentive and pre-attentive processes in Brooks' robots. Second, contrary to Brooks' ideas we have some sort of centralized representations in long term memory, albeit not of a traditional nature.

References

[Agre and Chapman 1987] Agre, P.E. and Chapman, D. (1987) "Pengi: an implementation of a theory of activity", AAAI'87, p268-272

[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. and Ramsey, A. (eds) Prospects for Artificial Intelligence, p229-238, IOS Press Amsterdam

[Botelho and Coelho 1995] Botelho, L.M. and Coelho, H. (1995) "A schema-associative model of memory", Proc. of the 4th 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", Proc. of the 4th International Workshop on Artificial Intelligence in Economics and Management (AIEM'96)

[Botelho and Coelho 1996b] Botelho, L.M. and Coelho, H. (1996) "Agents that rationalize their decisions", Submitted to the ECAI'96

[Botelho and Coelho 1996c] Botelho, L.M. and Coelho, H. (1996) "Learning by mood regulation", Submited to the Brazilian Symposium on Artificial Intelligence, SBIA96

[Botelho and Coelho 1996d] Botelho, L.M. and Coelho, H. (1996) "Emotion-based attention shift in autonomous agents", Submited to the ECAI'96 Workshop on Agent Theories, Architectures and Languages, ATAL96

[Bratman et al., 1988] Bratman, M.E., Israel, D. and Pollack, M.E. (1988) Plans and resource bounded practical reasoning, Computational Intelligence 4:349-355

[Brooks 1991a] Brooks, R. (1991) "Intelligence without representation", Artificial Intelligence, 47:139-159

[Brooks 1991b] Brooks, R. (1991) "Intelligence without reason", IJCAI'91, p569-595

[Clancey 1993] Clancey, W. (1993) "Situated action: a neuropsychological interpretation. Response to Vera and Simon", Cognitive Science 17:87-116

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

[Forgas 1994] Forgas, J.P. (1994) "The role of emotion in social judgments: an introductory review and an affect infusion model (AIM)", European Journal of Social Psychology, 24:1-24

[Georgeff and Ingrand 1989] Georgeff, M.P. and Ingrand, F.F. (1989) "Decision making in an embedded reasoning system", IJCAI'89, p972-978

[Kinny and Georgeff 1991] Kinny, D.N. and Georgeff, M.P. (1991) "Commitment and effectiveness of situated agents", IJCAI'91, p82-88

[Kirsh 1991] Kirsh, D. (1991) "Today the earwig, tomorrow man?", Artificial Intelligence, 47:161-184

[Laird and Bresler 1991] Laird, J.D. and Bresler (1991) "The process of emotional experience: a self-perception theory" in Clark, M. (ed) Review of Personality and Social Psychology, Sage, Berverly-Hills

[Lazarus 1984] Lazarus, R.S. (1984) "On the primacy of cognition", American Psychologist, 39:124-129

[Maes 1989] Maes, P. (1989) "The dynamics of action selection", IJCAI'89, p991-997

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

[Murphy and Zajonc 1993] Murphy, S.T. and Zajonc, R.B. (1993) "Affect, cognition and awareness: affective priming with optimal and suboptimal stimulus exposures", Journal of Personality an Social Psychology, 64:723-739

[O' Rorke and Ortony 1994] O' Rorke, P. and Ortony, A. (1994) "Explaining emotions", Cognitive Science, 18:283-323

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

[Pollack 1992] Pollack, M.E. (1992) The uses of plans, Artificial Intelligence 57:43-68

[Pollack and Ringuette 1990] Pollack, M.E. and Ringuette, M. (1990) "Introducing the TILEWORLD: experimentally evaluating agent architectures", AAAI'90, p183-189

[Roediger 1990] Roediger, H.L. (1990) "Implicit memory. Retention without remembering", American Psychologist, 49:1043-1056

[Ross 1977] Ross, L. (1977) The intuitive psychologist and his shortcomings: distortion in the attribution process, Advances in Experimental Social Psychology, 10:174-221

[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 1995] Sloman, A. (1995) "What sort of control system is able to have a personality?" To appear in the 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, New York

[Wright 1995] Wright, I.P. (1995) "Cognition and currency flow. Notes towards a circulation of value theory of emotions", unpublished document of the Cognitive Science Research Center of the University of Birmingham, UK (available with permission of the author: I.P.Wright@cs.bham.ac.uk)

[Zajonc 1980] Zajonc, R.B. (1980) "Feeling and thinking: preferences need no inferences", American Psychologist, 35:151-175

[Zajonc 1984] Zajonc, R.B. (1984) "On the primacy of affect", American Psychologist, 39:117-123