SC-EUNE - Surprise/Curiosity-based Exploration of UNcertain and UNknown Environments

Luís Macedo[1,2], Amílcar Cardoso[2]
[1] Instituto Superior de Engenharia de Coimbra, Quinta da Nora, 3031-601 Coimbra - Portugal
[2] CISUC-Centro de Informática e Sistemas / Departamento de Engenharia Informática da Universidade de Coimbra; Pinhal de Marrocos, 3030 Coimbra - Portugal
lmacedo@isec.pt; amilcar@dei.uc.pt

Abstract

This paper describes the architecture of SC-EUNE, a simulated artificial agent that explores uncertain and unknown environments guided by forms of surprise and/or curiosity. Besides describing the environment where it acts, we present the following components of its architecture: percepts, actions, goals, memory, emotions, drives and other motivations (currently confined to surprise and curiosity), and deliberative reasoning/decision-making. In addition, we discuss some experiments performed to assess the model of surprise implemented in SC-EUNE.

1 Introduction

Artificial Intelligence attempts to understand and build intelligent agents. Particularly, one of the approaches of Artificial Intelligence aims to understand and build artificial agents that act and think like humans (Russel and Norvig, 1995). Roughly speaking, agents accept percepts from the environment and generate actions. Their success depends heavily on the rationality of their actions, i.e., they perform well when they do the "right" action. Since different actions may lead to different states of world, in order to perform well, some kinds of agents make use of a mathematical function that maps a state of the world onto a real number - the utility value. States with higher utility values are preferred over states with lower utility values. Thus, in this kind of agents, decision-making is simply performed by selecting the action that leads to the state of the world with the highest utility value (Shafer and Pearl, 1990; Russel and Norvig, 1995).

Unfortunately, the real world is not crystal clear to the agent. Actually, agents almost never have access to the whole environment, mainly because of the incompleteness and incorrectness of their perceptual and understanding components. Actually, it is too much work to obtain all the information from a complex and dynamic world, and it is quite likely that the accessible information suffers distortions. Nevertheless, the agent should do always the "right thing", even when acting in those uncertain worlds. Probability theory provides the basis for treatment of agents that reason under uncertainty. The degree of expecting that an event X occurs in a random experiment is given by its probability P(X) (Shafer and Pearl, 1990).

When acting in those uncertain and unknown environments, although other agent's capabilities are very important (Etzioni and Weld, 1995), the learning capability is of extreme importance for agents. It provides them with knowledge that is crucial to improve future acts. This learning task may be guided by goals (Ram and Leake, 1995).

Another important attribute that agents should exhibit is personality, i.e., the capability of manifesting the attributes of a believable character such as emotion (Etzioni and Weld, 1995). Actually, considered by many authors as the principal motivational system, emotion is one of the sub-systems that compose personality (Izard, 1991). Another important sub-system is the drive system (also an important kind of the motivational system). Psychological and neuroscience research over the past decades suggests that emotions play a critical role in decision-making, action and performance, by influencing a variety of cognitive processes (e.g., attention, perception, planning, etc.). Actually, on the one hand, recent research in neuroscience (Damásio, 1994; LeDoux, 1996; Adolphs, 1996) supports the importance of emotions on reasoning and decision-making. For instance, results from recent studies of patients with lesions of the prefrontal cortex suggest an important role of emotions in decision-making (Damásio, 1994; Curchland, 1996; Bechara et al., 1997). According to Damásio's experiments, these patients are unable to make good decisions and had a strong impairment on their emotional assessment of situation, although pure cognitive abilities such as the ones measured by the traditional I.Q. rating remained unchanged. On the other hand, there are a few theories in psychology relating motivations (including drives and emotions) to action (Izard, 1977, 1991). For instance, in the specific case of emotions, as outlined by

Reisenzein (1996), within the context of the belief-desire theories of action (the dominant class of theories in today's motivation psychology) there have been proposals such as that emotions are action goals, that emotions are or include action tendencies, that emotions are or include goal-desires, and that emotions are mental states that generate goal-desires.

Although research in Artificial Intelligence has almost ignored this significant role of emotions on reasoning, several models has been proposed for emotions in general (some of them including surprise), focusing aspects such as: recognition of emotions (Picard, 1997), synthesis of emotions (Elliot, 1992; Bates, 1994; Blumberg, 1994; Maes, 1995; Reilly, 1996; Cañamero and Fredslund, 2000), synthesis of emotions and their influences decision-making and learning (Pfeifer, 1988; Kitano, 1995; Frijda, 1996; Barnes and Thagard, 1996; Velásquez, 1997; Scheutz et al., 2000). Some of these works also have addressed the role of drives and other motivations (including curiosity) on reasoning (e.g.: Velásquez, 1997).

Considered by many authors as a biologically fundamental emotion (e.g.: Izard, 1977, 1991; Ekman, 1992), surprise may play an important role in cognitive activities, especially in attention focus (Izard, 1977, 1991; Ortony and Partridge, 1987; Meyer et al., 1997; Reisenzein, 2000b), learning (Schank, 1982) and creativity (Boden, 1995; Macedo and Cardoso, 2001) (notice that some authors, like Ortony et al. (1988), do not consider surprise as an emotion). For instance, Meyer et al.'s experiments provide evidence indicating that surprising-eliciting events initiate a series of mental processes that begin with the appraisal of a cognised event as exceeding some threshold value of unexpectedness or schema discrepancy, continue with the interruption of ongoing mental information processing and the reallocation of processing resources to the event that has originated surprise, and culminate with an analysis and evaluation of that event plus immediate reactions to it and/or schema (beliefs) updating/revision. According to these authors, surprise has two main functions: informational and motivational. Thereby, it promotes both immediate adaptive actions to the surprise event and the prediction, control and effective dealings with future occurrences of the event. Moreover, there seems to be no doubt that creative products cause surprise and are to some extent constructed through a surprise-guided process (Macedo and Cardoso, 2001). Although defending that surprisingness and expectation failure are closely related concepts, Ortony and Partridge (1987) argue that in the real world there can be many situations in which one can be surprised by something one didn't expect without having to expect (consciously or unconsciously) something else. This model shares aspects with the one proposed by the research group of the University of Bielefeld, in Germany (e.g.: Meyer et al., 1997; Reisenzein, 2000a, 2000b), especially in that both defend that unexpected events elicit surprise. Implemented in a computer vision system, Peters' (1998) computational model of surprise roughly consists of a detection of unexpected movements. Other authors - e.g.: Velásquez (1997) - have also modelled surprise, although without the centrality given in the above ones.

A series of experiments (e.g.: Berlyne, 1950) have shown that in the absence of (or despite) known drives, humans, and in generally all animals, tend to explore and investigate their environment as well as seek stimulation. Curiosity is the psychological construct that has been closely related with this kind of behaviour. McDougall (1908) considered curiosity as an instinct and that it is instigated by novelty. Considering that curiosity was innate and that it could also be acquired, Berlyne, likewise McDougall, argued that novel stimulus elicits curiosity, which diminishes with continued exposure to the stimulus. Sharing similar ideas with Berlyne and McDougall, Shand (1914) defined curiosity as a primary emotion consisting of a simple impulse to know, controlling and sustaining the attention, and evoking the bodily movements that allow one to acquire information about an object. These approaches are closely related to the emotion concept of interest-excitement proposed by the differential emotions theory to account for exploration, adventure, problem solving, creativity and the acquisition of skills and competencies in the absence of known drives (Izard, 1991).

Moreover, as argued by Berlyne, in addition to novelty, other variables such as change, surprisingness, complexity, uncertainty, incongruity and conflict also determine this kind of behaviour related to exploration and investigation activities.

In this paper we describe SC-EUNE, a software agent whose goal is to explore an uncertain and unknown environment driven by forms of surprise and/or curiosity. Besides describing the environment where it acts, we present the following components of its architecture: percepts; actions; goals; memory; emotions, drives and other motivations; and, deliberative reasoning/decision-making.

The next section presents an overview of the architecture of SC-EUNE. The different component modules are described briefly in each subsection. Section 3 presents a session of SC-EUNE acting in an environment comprising buildings. In section 4 we made a discussion of some experimental results. Section 5 presents our further work. At last, some conclusions are made.

2 SC-EUNE's Architecture

SC-EUNE has the goal of exploring an uncertain and unknown environment comprising a variety of objects,

guided by forms of surprise and curiosity. Actually, besides desiring to know or be aware of all the objects belonging to the environment, SC-EUNE is able to feel the surprise and the curiosity that those objects cause to it. In fact, surprise and curiosity are the main goals that drive the exploring task of SC-EUNE: at a given time, among several objects available in the environment, it selects for study and analysis the one that causes to it more surprise and/or curiosity. This process is repeated until all the objects in the environment become known. Its behavior might be compared to an explorer seeking for emotions of surprise and new information. As many other agents, SC-EUNE has percepts, actions, goals, memory, emotions/drives units, and deliberative reasoning/decision-making (see Figure 1). The next subsections introduce each one of those components, as well as the environment where it acts.

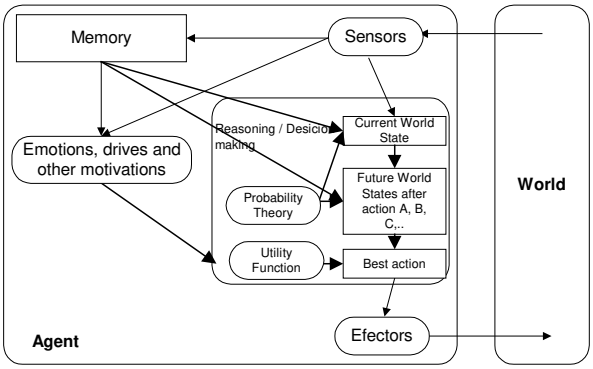

Figure 1 – Agent's architecture. The ovals represent processing modules while the rectangles represent information modules.

2.1 Environment

Previously defined by the user, the environment comprises a variety of objects located at specific positions. Represented by a graph (Macedo and Cardoso, 1998), each object comprises three distinct and fundamental components: *structure*, *function* and *behaviour* (Goel, 1992). The *structure*, the visible part of the object, corresponds to the shape of the object, which may comprise several other sub-objects. The *function* describes the role of the object in the environment. The *behavior* is related to the activity (actions and reactions) of the object in response to external or internal stimuli. Figure 2 shows an example of an environment where SC-EUNE can act. In this case, the objects are confined to buildings. The *structure* of the buildings comprises the shape (triangular, rectangular, etc.) of the roof, facade, door and windows. The possible *functions* may be: house, church, hotel, hospital, etc. Since those buildings are inanimate, they have no particular behavior, being simply described as static or inanimate. Figure 3 presents the graph-based representation and correspondent graphic representation of a building.

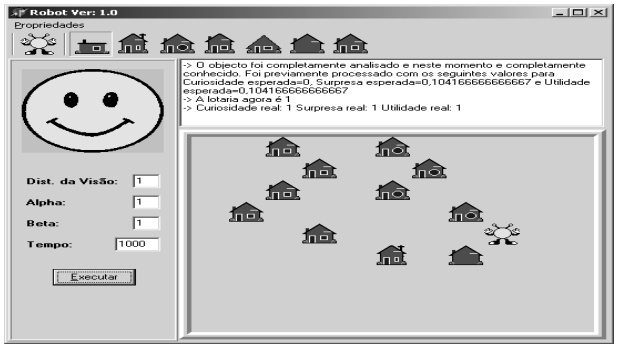

Figure 2- An example of an environment for SC-EUNE comprising buildings.

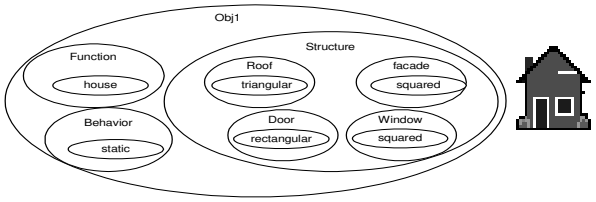

Figure 3 - Graph-based representations and respective graphic representation of the shape of a building, which is a house.

2.2 Perception

The perceptual system of the agent comprises two simulated sensors: an optic sensor and an infrared sensor. The information related to the *structure*, the *function*, and the *behavior* of the objects is collected from the environment through the optic sensor. The infrared sensor provides the distance of the objects. There is an user definable parameter for the range of the visual field. Objects out of that range are not visible by the agent. Notice that the *function* of the objects is not accessible (or can not be inferred from visual information) unless the agent is at the same place of the object.

2.3 Action

The main deliberative action executed by the agent is moving to a certain location in the environment, in which, unless there are no objects in the visual field, there is an object worthy of study and analysis. At a given time, this action corresponds to multiple actions as many existing different possible destination objects.

2.4 Goals

As we said above, the main goal of SC-EUNE is to explore the entire environment. Wanting to be at a given location or next to a given object may be considered as sub-goals.

2.5 Memory

As a knowledge-based agent, SC-EUNE stores all the information acquired through the sensors in its memory unit. The knowledge is of episodic kind: each object, stored as a case, is graph-based represented (as described above in section 2.1) in the episodic memory. Besides that, each case is associated with a number that expresses its absolute frequency. Figure 4 presents an example of the episodic memory of SC-EUNE after exploring an environment. Notice that at the beginning of an exploration the episodic memory is empty.

Field \ Case	C_1	C_2	C_3	C_4
Structure				
Function	House	House	School	Hotel
Behavior	Static	Static	Static	Static
Abs. Freq.	70	20	5	5

Figure 4 – Example of the episodic memory of SC-EUNE after exploring an environment.

In addition to the list of objects stored as cases, the agent also has a map of the environment in memory where it stores the location of the objects. This map is empty at the beginning, being continuously updated as new parts of the environment become known.

2.6 Emotions, drives and other motivations

At the current stage of development of the agent, this module is confined to surprise and curiosity. Nonetheless, other emotions, drives and other motivations will be included in this module in the future (see the section below about further work).

The objects of the environment may elicit surprise and curiosity in SC-EUNE.

The agent is almost continuously presented with an *input proposition* (Ortony and Partridge, 1987), which in the case of the environment described above corresponds to some information (visual or not) of a building (for instance, " a house with squared windows"). In response to this external stimulus, the surprise and curiosity unit outputs the intensity of these motivations, respectively.

In what concerns to surprise, together with Ortony and Partridge (1987), we think that it may result from three situations: (i) *active expectation failure*: resulting from a conflict or inconsistence between the *input proposition* and the *active prediction* or *expectation* (i.e., propositions explicitly represented in memory); (ii) *passive expectation failure* (or *assumption failure*): resulting from a conflict or inconsistence between the *input proposition* and what the agent knows or believes (*passive expectation or assumptions*), i.e., propositions that are not explicitly represented but that may be inferred easily; (iii) *unanticipated incongruities* or deviations from norms: resulting from a conflict or inconsistence between the input proposition and what, after the fact, may be judged to be normal or usual. Notice that in this last case, there are no *expectations* (*passive* or *active*) with which the *input proposition* might conflict.

In their cognitive-psychoevolutionary model, the research group of the University of Bielefeld (e.g.: Meyer et al., 1997) has defended similar ideas to those presented by Ortony and Partridge, namely that surprise (considered by them as an emotion) consists of the appraisal of unexpectedness. They suggest that surprise-eliciting events give rise to a series of the following mental processes: (i) appraisal of a cognised event as exceeding some threshold value of unexpectedness (schema-discrepancy) - according to Reisenzein (in press), there is a comparator, the appraisal function, that computes the degree of discrepancy between new and old beliefs or schemas; (ii) interruption of ongoing information processing and reallocation of processing resources to the investigation of the unexpected event; (iii) analysis/evaluation of that event; (iv) possibly, immediate reactions to that event and/or updating or revision of the "old" schemas or beliefs.

Let us now describe how the intensity of surprise is computed. There is experimental evidence supporting that the intensity of felt surprise increases monotonically, and is closely correlated with the degree of unexpectedness - see (Reisenzein, 2000b) for a review of these experiments. This means unexpectedness is the proximate cognitive appraisal cause of the surprise experience. Considering this evidence we propose that the surprise felt by an agent *Agt* elicited by an object Obj_k is given by the degree of unexpectedness of Obj_k, considering the set of objects present in the memory of the agent *Agt:*

$$SURPRISE(Agt, Obj_k) = DegreeOfUnexpectedness(Obj_k, Agt(Mem))$$

According to probability Theory (e.g.: Shafer and Pearl, 1990), the degree of expecting that an event X occurs is given by its probability P(X), and then the improbability of X, denoted by 1-P(X), defines the degree of not expecting X. Thus, the surprise function is as follows:

$$SURPRISE(Agt, Obj_k) = 1 - P(Obj_k)$$

Although other probabilistic methods might be used to compute P(X), in case of objects comprising several components we may compute the probability of the whole object Obj_k computing the mean of the conditional probabilities of their *n* constituent parts, which are individually computed using the Bayes's formula (Shafer and Pearl, 1990) (notice that each one of those

conditional probabilities individually gives the degree of unexpectedness of a specific piece of the object, given as evidence the rest of the object):

$$P(Obj_k^l) = \frac{\sum_{l=1}^{n} P(Obj_k^l \mid Obj_k^1, Obj_k^2, ..., Obj_k^{l-1}, Obj_k^{l+1}, ..., Obj_k^n)}{n}$$

It is worth noticing that there is a correspondent facial expression for the intensity of surprise generated (Ekman, 1992).

As defined by McDougall, Berlyne and Shand, curiosity is elicited by novelty. Actually, being defined as the desire to know or learn an object that arouses interest, as by being novel or extraordinary, it means that novel and possibly unexpected objects stimulate actions to acquire knowledge of those objects. These actions usually comprise, firstly, the focus of the senses on the unknown object. For instance, there is no doubt that humans usually focus their eyes in the new objects of an environment. Actually, when faced with a set of objects, they are more attracted by new objects. Objects that are familiar to the agent do not attract them as new ones do, at least for a few moments. Thus, according to this consensual definition, the curiosity induced in an agent *Agt* by an object Obj_k is given by the novelty or difference of Obj_k relatively to the set of objects present in the memory of *Agt*:

$$CURIOSITY(Agt, Obj_k) = DIFFERENCE(Obj_k, Agt(Mem))$$

Roughly speaking, the measure of difference is described as follows: relying on error correcting code theory (Hamming, 1950), the function computes the distance between two graph-based represented objects, counting the minimal number of changes (insertions and deletions of nodes and edges) required to transform one graph-based represented object into another (e.g., Messmer and Bunke, 1998). To compute the difference of a given object relatively to a set of objects, we apply the above procedure to each pair of objects formed by the given object and an object from the set of objects. The minimum of those differences is the difference of the given object relatively to the given set of objects.

2.7 Deliberative Reasoning/Decision-making

As we said above, SC-EUNE is always moving from location to location, from object to another object, exploring the environment. At a given time, SC-EUNE has to decide the location where to move. Among the available actions such as *MoveTo*(*Obj1*), *MoveTo*(*Obj2*), *MoveTo*(*LocationXY*),…, the deliberative reasoning/decision-making module selects the one that seems to be the most profitable in terms of surprise and/or curiosity. This means, the actions are selected taking into account their results in what concerns to the intensity of surprise and/or curiosity caused to the agent. The results (new world state) of the available actions may be described, for instance, by *CloseTo*(*Obj1*), *CloseTo*(*Obj2*),…, or *At*(*LocationXY*), etc. As a classical decision-maker, SC-EUNE's deliberative reasoning/decision-making component makes use of an Utility Function (Shafer and Pearl, 1990; Russel and Norvig, 1995). This Utility Function maps each one of those possible states or results into a real number. Since this number expresses the intensity of the surprise and/or of the curiosity felt by the agent at those states, the Utility Function of a state *S* (described for instance by *CloseTo*(Obj_k), etc.) depends on another two Utility Functions, which are the Utility Function of surprise (denoted by *Us*) and the Utility Function of curiosity (denoted by *Uc*), as follows:

$$U(S) = \frac{\alpha \times Uc(S) + \beta \times Us(S)}{\alpha + \beta}$$

, where α and β are the weights of the Utility Functions of surprise and of curiosity, respectively. These functions make use of the functions of surprise and curiosity that belongs to the emotions, drives and other motivations module (described in section 2.6) as follows:

$$Us(S) = Us(CloseTo(Obj_k)) = SURPRISE(Obj_k, Agt(Mem))$$

$$Uc(S) = Uc(CloseTo(Obj_k)) = CURIOSITY(Obj_k, Agt(Mem))$$

Thus, given a set of new states S_1, S_2, …S_n, resulting from actions A_1, A_2, …, A_n, respectively, the agent applies the above function to each one of them and selects the action that maximizes it. This means the agent is obeying to the Maximum Utility Principle. However, it might happen that an action A_i may lead to one of a set of possible states (there is no certainty of which one, but a prediction). In SC-EUNE's environment, this happens when an object is incompletely known (i.e., the *function* is lacking), which occurs when the SC-EUNE is looking at an object which is not at the same place of it. In this case, moving to that object is a lottery (Shafer and Pearl, 1990; Russel and Norvig, 1995) as follows:

$$Lottery(MoveTo(Obj)) = \\ = [p_1, CloseTo(Obj_1); p_2, CloseTo(Obj_2); ...; p_n, CloseTo(Obj_n)]$$

, where p_i is the probability of *Obj* be Obj_i.

For each lottery of this kind, the agent applies the Expected Utility Function (denoted by *EU*):

$$EU(p_1, S_1; p_2, S_2; ...; p_n, S_n) = \sum_i p_i \times U(S_i)$$

, where S_i denotes the i^{th} possible result or state of the lottery. Notice that in this case, $S_i = CloseTo(Obj_i)$. Given a set of actions, each one giving rise to a lottery, the agent, as we said above, applies the above Expected Utility Function to each one of them and selects the one that maximizes the function. This way the agent is applying the Maximum Expected Utility Principle. Notice that if there is only one possible state S_1 (when $n = 1$ in the above equation) for a given action then p_1=1 and then,

$$EU(1,S_1)=1\times U(S_1)=U(S_1)$$

Let us now summarize the behavior of SC-EUNE. At a given time, two main situations may happen: there are objects in the visual field, or there are no objects in the visual field. In the former situation, SC-EUNE applies the Expected Utility Function to each one of the objects present in the visual field. Then SC-EUNE selects the action of moving towards the object that maximizes that function. In the second situation, if the agent reminds objects that have not yet been visited (these incompletely known objects are stored in the episodic memory of the agent and marked as not yet studied or visited) then it applies to them the Expected Utility Function in a similar procedure as to the one described above for objects in the visual field (although in this case the objects are not in the visual field, they are imaged in the memory of the agent). If there are no incompletely known objects in memory nor in the visual field, then the agent decides to move one location forward so that it can find more worthy of study objects.

3 A Session

Let us illustrate the behaviour of SC-EUNE as described above.

Consider the scene depicted in Figure 5, in which SC-EUNE in an environment with several buildings. Considering the current position of SC-EUNE as depicted in Figure 5, *ObjA* (a static house with a rectangular door, a squared facade, squared windows and an isosceles triangular roof) and *ObjB* (a static house with a rectangular door, a squared facade, circular windows and an isosceles triangular roof) are in the visual field of SC-EUNE (the visual field was set to 3 distance units[1]). The starting memory of SC-EUNE is the one described in Figure 4.

[1] These units are of abstract kind, having no relation to any known distance units such as metres, centimetres, etc.

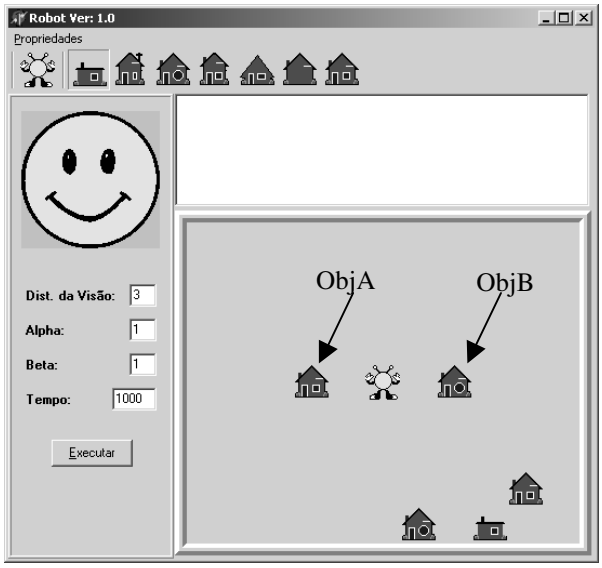

Figure 5 – SC-EUNE in an environment with several buildings.

SC-EUNE starts by computing the lotteries of moving to these two objects. Notice that there are three possibilities for the function of those buildings: *house*, *school* and *hotel*. Thus moving to *ObjA* may correspond to moving to a *house* ($ObjA_1$), a *school* ($ObjA_2$) or a *hotel* ($ObjA_3$):

$$Lottery(MoveTo(ObjA))=$$
$$=\left[0.89,CloseTo(ObjA_1);0.058,CloseTo(ObjA_2);0.052,CloseTo(ObjA_3)\right]$$

In an analogous way, moving to *ObjB* may correspond to moving to a *house* ($ObjB_1$), a *school* ($ObjB_2$) or a hotel ($ObjB_3$):

$$Lottery(MoveTo(ObjB))=$$
$$=\left[0.9,CloseTo(ObjB_1);0.05,CloseTo(ObjB_2);0.05,CloseTo(ObjB_3)\right]$$

Applying the Expected Utility Function to each one of those lotteries, yields,

$$EU(0.89,CloseTo(ObjA_1);0.058,CloseTo(ObjA_2);0.052,CloseTo(ObjA_3))=$$
$$=0.89\times U(CloseTo(ObjA_1))+0.058\times U(CloseTo(ObjA_2))+$$
$$+0.052\times U(CloseTo(ObjA_3))=0.89\times 0.123+0.058\times 0.228+0.052\times 0.258=$$
$$=0.136$$

$$EU(0.9,CloseTo(ObjB_1);0.05,CloseTo(ObjB_2);0.05,CloseTo(ObjB_3))=$$
$$=0.9\times U(CloseTo(ObjB_1))+0.05\times U(CloseTo(ObjB_2))+$$
$$+0.05\times U(CloseTo(ObjB_3))=0.9\times 0.183+0.05\times 0.342+0.05\times 0.342=0.199$$

As described above, to compute each one the elements $U(CloseTo(Obj_i))$ in the above expressions, SC-EUNE makes use of the Utility Functions of surprise and curiosity (*Us* and *Uc*). Let us illustrate this computation for the case of $U(CloseTo(ObjA_1))$. As defined above, this is given by,

$$U(CloseTo(ObjA_1)) = \frac{\alpha \times Uc(CloseTo(ObjA_1)) + \beta \times Us(CloseTo(ObjA_1))}{\alpha + \beta}$$

Computing separately $Us(CloseTo(ObjA_1))$.and $Uc(CloseTo(ObjA_1))$, comes,

$$Us(CloseTo(ObjA_1)) = SURPRISE(ObjA_1, Agt(Mem)) =$$
$$= 1 - P(ObjA_1) = 1 - 0.92 = 0.08$$

$$Uc(CloseTo(ObjA_1)) = CURIOSITY(ObjA_1, Agt(Mem)) =$$
$$= DIFFERENCE(ObjA_1, Agt(Mem)) = 0.166$$

Considering $\alpha = \beta = 1$, results,

$$U(CloseTo(ObjA_1)) = \frac{1 \times 0.166 + 1 \times 0.08}{1+1} = 0.123$$

Considering the Expected Utility Values computed above, SC-EUNE selects moving to *ObjB*. Actually, the Expected Utility Value of this object (0.199) is greater than that of *ObjA* (0.136). After becoming aware that *ObjB* is a *house*, i.e., that *ObjB* is $ObjB_1$, the Utility Value of this object (now a real value rather than an expected value) is computed again, but now regarding to the complete information of the object that is now completely available. In this case, the values of surprise and curiosity are 0.2 (20%) and 0.166 (16,6%), respectively, and the Utility Value is 0.183. Since *ObjB* is no yet in memory, SC-EUNE stores it in the episodic memory with the absolute frequency of 1.

At this point, SC-EUNE has no more objects in the visual field (notice that *ObjA* is now at 6 units of distance). However, SC-EUNE has marked in its episodic memory *ObjA* as not yet visited, and then it can compute again its Expected Utility Value. This computation has to be done again because there were changes in the episodic memory of the agent that may have implications on that computation. Since there are no more alternative buildings to visit, SC-EUNE selects the action of moving to *ObjA*. After getting the information that it is a *house*, SC-EUNE updates the absolute frequency of this object to 71 (notice that it already has in memory 70 buildings similar to this one).

At this point, there are neither more objects in the visual field nor more unvisited objects in the episodic memory. Therefore, SC-EUNE moves one location forward. This action is maintained until another unvisited object appears in its visual field.

4 Discussion of some Experiments

Although all the experimental evidence reported above supports our models of surprise and curiosity, experimental tests are required in order to assess them with more accuracy.

In the particular case of surprise we have performed experiments to test mainly whether the intensity values rated by the artificial agent match the ones rated by humans under similar circumstances. The experiments were carried out in two domains: the domain of buildings (described in this paper) and in an abstract domain with hedonically neutral events (see Stiensmeier-Pelsteret al. (1995) for a similar experiment with humans). Those experiments comprise the presentation of several quiz items to participants, who were asked to give *active* or *passive expectations* for a particular element of an event depending on whether or not this element was shown to them. In addition, the intensity of felt surprise regarding to that element (after showing it in the case it was hidden) and to the whole event was asked. The results of those experiments vary depending on both the amount of knowledge stored in the episodic memory of the artificial agent and on the domain considered. However, best results were achieved with the hedonically neutral domain, in which was obtained an average difference of the order of 6% between the values given by the humans and by the artificial agent. We have also concluded that the surprise induced by a same building in different times of the exploratory task may change. This is explained by the changes performed in the episodic memory of the agent during that exploratory task. Actually, as new buildings are perceived and visited, the absolute frequencies of the episodes change, having implications on the surprise function that depends on those frequencies. Recent experiments with children and adults have provided evidence indicating that in humans the intensity of felt surprise depends on the amount of knowledge one has (see Schutzwohl and Reisenzein, 1999).

5 Further Work

In the more general agent on which we are working we plan to integrate other emotions, drives and other motivations in the Utility Function, which may look as follows:

$$U(S) = \frac{\alpha_1 \times Ue1(S) + \alpha_2 \times Ue2(S) + ... + \beta_1 \times Ud1(S) + ... + \delta_1 \times Uo1(S) + ...}{\sum_i \alpha_i + \sum_j \beta_j + \sum_k \delta_k}$$

where α_i, β_j and δ_k are the weights of emotion e_i, drive d_j and other motivation o_k, respectively, with $i, j, k \in \aleph$, and S is the state of the world resulting from an available action.

Negative and positive emotions (Izard, 1977) play the role of decreasing and increasing, respectively, the utility values. Since emotions may interfere with each other and can amplify, attenuate, or inhibit drive systems, a possible approach to deal with this may be to consider the weights of those dependent emotions, drives and other motivations as functions of other emo-

tions. Another approach may be to consider the Utility Functions themselves as functions of other emotions.

By using the above Utility Function we are considering that emotions are action goals (Reisenzein, 1996). Actually, the goal of an agent with that Utility Function is maximization of positive feelings and minimization of negative feelings, i.e., the goal is to feel pleasant emotions and avoid unpleasant ones. This approach is to some extent related to the *somatic marker hypothesis* of Damásio (1994). However, other approaches such as considering emotions as mental states that generate goal-desires might be considered. Furthermore, experiments may be performed to compare these approaches.

Notice that the weights used in the Utility Function allow us to model different agents and also agents with disorders in the emotional system (e.g.: agents with no fear). Actually, we may build agents that use the above Utility Function with different weights. Moreover, the consideration of different thresholds of activation in the Utility Functions of emotions, drives and other motivations may contribute also to modelling different agents.

Additional experiments should be performed to assess these ideas and particularly the role played by those emotions, drives and other motivations not only in the exploratory task but also in other activities such as planning and creativity.

6 Conclusions

We have presented an architecture for an artificial agent that explores uncertain and unknown environments guided by human forms of surprise and/or curiosity. Besides describing the environment where it acts, we present the following components of its architecture: percepts, actions, goals, memory, emotions, drives and other motivations, and deliberative reasoning/decision-making.

References

R. Adolphs, R. et al. Neuropsychological Approaches to Reasoning and Decision-Making. In: A. Damásio et al., (Eds.), *Neurobiology of Decision-Making*. Berlin: Springer-Verlag, 1996.

A. Barnes and P. Thagard. Emotional Decisions. In: *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*. Erlbaum 426-429, 1996.

J. Bates. The Role of Emotion in Believable Agents. *Communications of the ACM*, 37(7):122-125, 1994.

A. Bechara, H. Damásio, D. Tranel and A. Damásio. Deciding Advantageously Before Knowing the Advantageous Strategy. *Science* 275:1293-1295, 1997.

D. Berlyne. Novelty and curiosity as determinants of exploratory behavior. *British journal of Psychology*, 41, 68-80, 1950.

B. Blumberg. Action-Selection in Hamsterdam: Lessons from Ethology. In: *Proceedings of* SAB'94, 108-117, Brighton, England: MIT Press, 1994.

M. Boden, M. *Creativity and unpredictability*. *SEHR*, 4, 2, 1995.

D. Cañamero and J. Fredslund. How does it feel? Emotional Interaction with a Humanoid LEGO Robot. In: K. Dautenhahn (ed.), *Socially Intelligent Agents: The Human in the Loop.* Papers from the AAAI 2000 Fall Symposium, pp. 23-28. Menlo Park, CA: AAAI Press, 2000.

P. Churchland. Feeling Reasons. In A. Damásio et al., (Eds.), *Neurobiology of Decision-Making*. Berlin: Springer-Verlag, 1996.

A. Damásio. *Descartes'error, Emotion Reason and the Human Brain*. Grosset/Putnam books, 1994.

P. Ekman. An Argument for Basic Emotions. In: Stein, N. L., and Oatley, K. eds. Basic Emotions, 169-200. Hove, UK: Lawrence Erlbaum, 1992.

C. Elliott. The Affective Reasoner: A Process Model of Emotions in a Multi-agent System. Ph.D. Dissertation, Northwestern University. The Institute for the Learning Sciences, Technical Report No. 32, 1992.

O. Etzioni and D. Weld. Intelligent Agents on the Internet: Fact, Fiction and Forecast. *IEEE Expert*, 10(4), 44-49, 1995.

N. Frijda. *The Emotions*. Cambridge, UK: Cambridge University Press, 1986.

A. Goel. Representation of Design Functions in Experience-Based Design. In: Brown, D., Walderon, M., Yosnikawa, H. (Eds.): *Intelligent Computer Aided Design*. Elsevier Science, 1992.

R. Hamming. Error Detecting and Error Correcting Codes. *The Bell System Technical Journal*, Vol. XXVI, 2, 147-60, 1950.

C. Izard. *The Psychologu of Emotions.* Plenum Press, New York, 1991.

C. Izard. *Human Emotions.* Plenum Press, New York, 1977.

H. Kitano. A Model for Hormonal Modulation of Learning. In: *Proceedings of IJCAI'95*, Montreal, 1995.

J. LeDoux. *The Emotional Brain*. New York, Simon and Schuster, 1996.

L. Macedo and A. Cardoso. Creativity and Surprise. *Proceedings of the AISB'01 Symposium on Artificial Intelligence and Creativity in Arts and Science*. 2001.

L. Macedo and A. Cardoso. Nested Graph-Structured Representations for Cases. *Proceedings of the 4th European Workshop on Case-Based Reasoning* (pp. 1-12). Berlin: Springer Verlag, 1998.

P. Maes. Artificial Life meets Entertainment: Lifelike Autonomous Agents. *Communications of the ACM. Special Issue on Novel Applications of AI*, 1995.

W. McDougall. *An introduction to social psychology*. London: Methuen, 1908.

B. Messmer and H. Bunke. A new algorithm for error-tolerant subgraph isomorphism detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 (5), 1998.

W. Meyer, R. Reisenzein and A. Schutzwohl. Towards a process analysis of emotions: The case of surprise. *Motivation and Emotion*, 21, 251-274, 1997.

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

A. Ortony and D. Partridge. Surprisingness and Expectation Failure: What's the Difference?. *Proceedings of the 10th International Joint Conference on Artificial Intelligence* (pp. 106-8). Los Altos, CA: Morgan Kaufmann, 1987.

M. Peters. Towards Artificial Forms of Intelligence, Creativity, and Surprise. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, pp. 836--841, Madison, Wisconsin, 1998.

R. Pfeifer. Artificial Intelligence Models of Emotion. In: V. Hamilton, G. Bower, and N. Frijda (Eds.). *Cognite Perspectives of Emotion and Motivation*, 287-320. Netherlands: Kluwer, 1988.

R. Picard. *Affective Computing*. MIT Press, 1997.

A. Ram and D. Leake. *Goal-driven learning*. Cambridge, MA: MIT Press, 1995.

S. Reilly. Believable Social and Emotional Agents. Technical Report, CMU-CS-96-138, School of Computer Science, Carnegie Mellon University, 1996.

R. Reisenzein. A theory of emotions as Metarepresentational States of Mind. *Personality and Social Psychology Reviews*, in press.

R. Reisenzein. Exploring the Strength of Association between the Components of Emotion Syndromes: The Case of Surprise. *Cognition and Emotion*, 1, 1-38, 2000a.

R. Reisenzein. The subjective experience of surprise. In H. Bless and J. Forgas (Eds.), *The message within: The role of subjective experience in social cognition and behavior*. Philadelphia, PA: Psychology Press, 2000b.

R. Reisenzein. Emotional Action Generation. In W. Battmann and S. Dutke (Eds.), *Processes of the molar regulation of behavior*. Lengerich: Pabst Science Publishers, 1996.

S. Russel and P. Norvig. *Artificial Intelligence - A Modern Approach*. Prentice Hall, 1995.

R. Schank. *Dynamic Memory*. NJ: Cambridge University Press, 1982.

G. Shafer and J. Pearl (Eds.). *Readings in Uncertain Reasoning*. Palo Alto, CA: Morgan Kaufmann, 1990.

A. Shand. *The foundations of character*. London: Macmillan, 1914.

M. Scheutz, A. Sloman and B. Logan. Emotional States and Realistic Agent Behaviour. In *Proceedings GAME-ON 2000*, 2000.

A. Shutzwohl and R. Reisenzein. Children's and Adults' Reactions to a Schema-discrepant Event: A Developmental Analysis of Surprise. *International journal of Behavioral Development*, 23(1), 37-62, 1999.

J. Stiensmeier-Pelster, A. Martini and R. Reisenzein. The role of surprise in the attribution process. *International journal of Behavioral Development*, 23(1), 37-62, 1995.

J. Vélasquez. Modeling Emotions and Other Motivations in Synthetic Agents. In: *Proceedings of AAAI-97*, 1997.