# CREATING AN EMOTIONAL SPACE FOR ARTIFICIAL BEINGS

Penny Baillie
Mark Toleman
Department of Mathematics and Computing
University of Southern Qld.
Toowoomba 4350. AUSTRALIA
penny.baillie@usq.edu.au

KEYWORDS
Intelligent, Autonomous Agents, Emotions, Affective Computing.

## ABSTRACT

Interacting intelligently within a temporally dynamic environment calls for truly adaptive performances from artificial beings. One mechanism currently being explored to produce an intuitive-like behaviour in AI applications is that of emotion. Several popular psychological theories of emotion have been the basis for a number of artificial agent models that are capable of synthesizing emotions. However, whilst these models have proved successful, they have been built on theories that were not conceived for use in computer applications and thus lack or do not explain clearly the internal variables and mechanisms needed for adaptive applications. This paper examines another psychological theory built on discrete evaluations of cognitive appraisal dimensions that overcomes this problem and demonstrates an artificial agent that has had this theory integrated into its architecture.

## INTRODUCTION

Although emotions have traditionally been seen as irrational interruptions to an otherwise logical state of being, there has currently been a resurgence of interest in studying their generation and effects. Modern studies have shown that emotions play an important part in rational behaviour. Damasio in (Damasio 1994) describes patients with frontal-lobe disorders in the brain that impede their ability to generate emotional responses. This impairment drastically impacts on the patient's aptitude to make rational decisions due to this lack of *somatic markers*. In healthy people, these markers associate positive and negative feelings with decision-making processes. One symptom of this disorder presents itself when a patient is asked to make a simple decision. Often the subject will become engrossed in a continuous and exhaustive search for a solution. An activity that is relatively simple for a healthy human, such as scheduling a meeting, becomes a plethora of possibilities, where no solution is apparent for a Damasio's patient.

Interestingly enough, this condition makes for a clea example of Hanley's philosophy (Hanley 1997) that withou emotion we could be endlessly suspended in a state o indecision if it were not for emotions. Furthermore, emotion have been found to bring rationality to chaotic behaviour i humans. They affect our reasoning and decision-making abilities. If one of the ultimate purposes of developing artificial intelligences is to mimic our intelligent behaviour then surely a machine devoid of emotional mechanisms wil never achieve that goal.

It may seem inconceivable that we would have the need t develop an emotional machine but with the technologica advancement of AI into vast domains, it may not be fanciful to think that the integration of emotion theory with AI would be advantageous. As we become more dependent on machines making choices for us, it could be comforting to know that they are making human like decisions with a 'digital heart'.

This paper presents our recent research with an artificial agent testbed and architecture designed to synthesize emotional states and to use this mechanism for affective decision making.

## ATTITUDES

Today's lifestyle means that we are constantly bombarded with decisions. Very rarely do these decisions involve right and wrong answers. Often we have to choose from a number of solutions that would all suit our purpose equally as well. For example, a person might decide to purchase a car. Having chosen the make and model, the final decision may be, *do I take the red one or the green one?* In this case, the colour will not affect the ability of the car to fulfill the requirements that it is being purchased for. So we make a decision based on personal preference. This disposition of feeling positively or negatively towards something is referred to, by psychologists, as attitude (Fishbein and Ajzen 1975).

Whatever the collection of past experiences that makes us prefer green cars to red cars, it defines who we are as individuals, our likes and dislikes and determines our behaviour. If we all held the same attitudes we would all be driving the same car, eating at the same restaurants and using the same washing power. But this is not the case. Ajzen and Fishbein as noted by Petty and Cacioppo in (Petty and

Cacioppo 1996) conclude that behaviour can be viewed as a collection of attitudes toward four key elements. They are:

- the action being performed;
- the target or targets that are the object of the action;
- the context of the action, for example, where it is being performed; and,
- the temporal alignment of the action, for example, the time of day or month.

By examining these elements we can form a conceptual understanding of a situation involving rational human behaviour and use these concepts to predict future behaviours by attitudinal assessment. The set of these elements defines, what we have termed, the *event space*.

The event space $E$, can be defined as:

$$E = \{a, o, c, t\} \qquad (1)$$

where $a$ is the set of actions that relate to the event, $o$ is the set of objects involved or affected by $a$, $c$ is the context or conditions in which $a$ is taking place or being performed and $t$ is the temporal component of $a$. Each of these elements may vary along a dimension of explicitness. For example, the temporal element could refer to a general time such as *2 o'clock* or it could be more explicit such as *2 p.m. on Friday the 25th of April*. At the most exact level of defining an event space, a person will intend to perform a specific action, with or towards a certain object, in a particular context or situation at an exact point in time.

By using a relevant collection of a person's attitudes toward the elements in past episodes of event spaces, it is possible to closely predict the behaviour of that person in future event spaces (Baillie (Ray), Toleman et al. 2000). This method of predicting behaviour from past behaviours is otherwise known as *the theory of reasoned action (Fishbein and Ajzen 1975)*. In this theory, an attitude towards an object, $A$ is expressed as:

$$A = \sum_{i=1}^{n} W_i V_i \qquad (2)$$

where $W_i$ is the weighting of blame or praise associated with the object being instrumental in the failure or success of a situation $i$. $V_i$ is the degree of satisfaction or dissatisfaction obtained from the situation and n is the number of times that an identical situation has involved the use of the object.

Ajzen and Fishbein (Fishbein and Ajzen 1975) also refer to a subset of attitudes called the *subjective norm*. This is a motivation to comply with society. The subjective norm is a person's belief that important reference groups or individuals endorse or reject particular types of behaviour, or

$$SN_B = \sum_{j=1}^{p} b_j m_j \qquad (3)$$

*where* $b_j$ is the set of $p$ number of beliefs about a behavi $B$, that a group of individuals $j$ think are acceptable unacceptable and $m$ is the motivation to comply with $b_j$. applying these formula to the four key elements of the ev space (mentioned above) we can find a summative atti towards a situation.

Before we can examine how attitude affects behaviour need to understand the structure of behavioural motivators a goals.

## APPLYING ATTITUDE TO GOAL SELECTION

Human behaviour is goal orientated (Baillie (Ray) a Lukose 1999). Koestler (Koestler 1967), to separate m from machine, makes the observation that the hum organism is not merely a mechanical device, but it reacts to ever-changing world and how it reacts to the world is bas on its goals at the time. How these goals are selected is bas on a person's individual needs and wants in a given situati (Baillie (Ray), Lukose et al. 2000). These needs and wants a determined by what we as humans believe our goals to b what triggers us to perform acts to fulfill these goals and wh those acts are. We as individuals, through our experienc have different ideas on what our goals are and how to achie them. This prompts us to act diversely when compared others yet to act rationally in our reasoning and decisi making. The old adage, "there is more than one way to skir cat", is all too true when comparing behaviour among humans. . Many people all with the same goal will act different ways to achieve it. All actions may be as sound a successful as the next and it is this that illustrates how we a affected by past experiences and ,attitudes.

Picard believes that emotions are an integral part and natural progression for artificial intelligence (Picard 199 She states that, *'the inability of today's computers recognize, express, and have emotions severely limits the ability to act intelligently'*. Understanding emotion generation and developing a reasonable empirical model needed to discover the potential of affective reasoning a decision-making in AI.

In a previous paper (Baillie, Toleman et al. 2000), w discussed an artificial agent that makes a decision by selectin a solution that it *likes* the most. This *liking* is the attitud assigned to a past event space and its elements based on th agent's success or failure during that event. For example, the agent was to attempt a task that failed, its attitude toward the elements of that event space would decrease making it le likely to put itself in that situation again. The agent can mak decisions on which solution to choose in the future b building event spaces for all its choices and calculating it attitude towards them using its attitudes about elementar items and beliefs about subjective norms.

An example of this type of decision making can be seen i Figure 1. In this simulation there are 2500 cells that represen 2500 individuals. At each time iteration, each individua makes a decision among five choices all of optimal outcome.

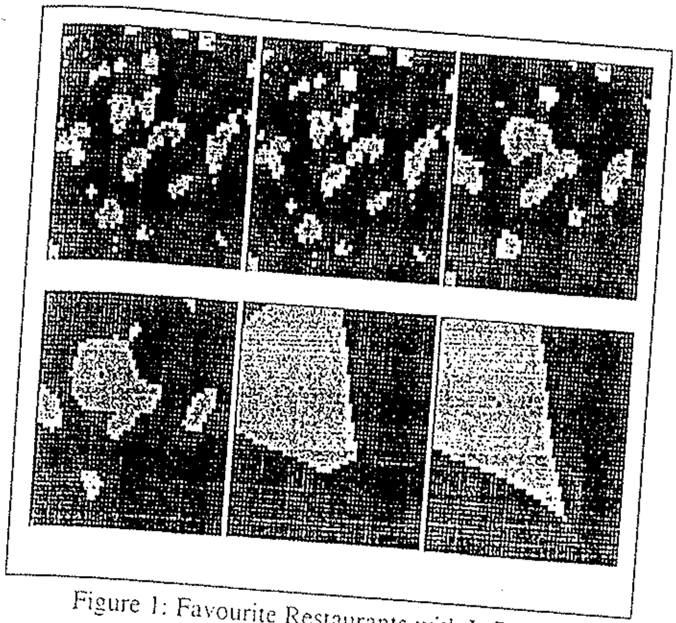


Figure 1: Favourite Restaurants with Influential Individuals

To visualize this, consider a population that has a choice of dining at one of five different restaurants. Initially none of the population has eaten at any of the restaurants and therefore they have no preconceived attitudes about them. On each iteration an individual selects a restaurant based on how it *feels* about it. On the first iteration individuals have neutral attitudes about each restaurant. Therefore, the individuals each pick one at random. Having chosen a restaurant to eat at, there is a chance (we set this to 10%) that the individual will have a bad experience. If this occurs the attitude held by the individual about the restaurant decreases otherwise its attitudinal beliefs about the restaurant are reconfirmed and its attitude improves. As intentions are not only governed by an individual's attitudes, before each individual makes a choice of which restaurant to eat at, it gathers together the attitudes about each restaurant from each of the individuals in the neighbouring cells and uses these when deciding.

We ran the simulation with several *prominent* individuals that had a stronger influence over the surrounding population. The shade of the cells in Figure 1 represents the favourite restaurant and the black cells represent the prominent individuals. Not unexpectedly, from random beginnings the individuals in close vicinity formulate the same attitudes about their favourite restaurant and groups of individuals that prefer the same restaurant begin to gather around the prominent individuals.

While this is an affective simulation, it is a very simplistic view of human decision-making. More realistically we tend to select options that make us *feel good* or return us to a state of well being (Koestler 1967). Often these choices may not be *liked*. For example, having a tooth pulled out to alleviate pain may not be a choice that is liked but the benefit outweighs the unpleasantness. Not only should we be assessing a solution based on its pleasantness but also on a number of other factors and how the assessment of these factors produces an emotional response.

## BUILDING COMPLEX EMOTIONS WITH COMPLEX ATTITUDES

Evaluating an event as *liked* or *disliked* gives very littl insight into the emotional state that the event may invoke. Fo example, *happiness* is a liked emotion as is *pride*. Differentiating between emotions requires more than a simple pleasantness factor.

To implement a system in our agents that can distinguish between emotional states, we have chosen the Smith and Ellsworth six dimensional model of emotion. This model (Smith and Ellsworth 1985) proposes 6 cognitive appraisal dimensions by which to differentiate emotional experiences. Pleasantness, which we mentioned above, is one of them. Their experimentation and analysis in (Smith and Ellsworth 1985), identifies 6 orthogonal dimensions: *pleasantness, anticipated effort, certainty, attentional activity, responsibility* and *control*, across fifteen emotions: *happiness, sadness, anger, boredom, challenge, hope, fear, interest, contempt, disgust, frustration, surprise, pride, shame* and *guilt*.

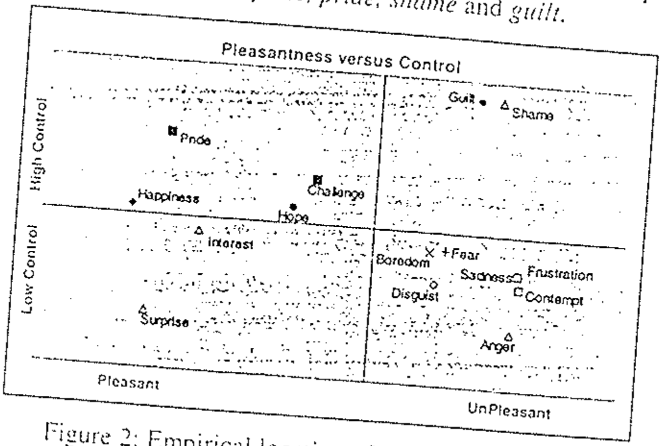


Figure 2: Empirical location of emotional states with respect to the pleasantness and control dimensions.

Figure 2 displays two of the orthogonal dimensions and where the fifteen emotional states are positioned with respect to them. Given this model we can assess event spaces by the emotional reaction that they evoke. Rather than appraising an event on its pleasantness (like or dislike) we can also make five other cognitive appraisals and experiment with an agent that can be made to gravitate toward any one or combination of fifteen emotions. With these values we can now build an *affective space* for an individual or artificial intelligence.

The affective space, *A*, can be defined as a 6 dimensional space occupied by points that ascertain the location of fifteen pure emotions. Each emotion is a point, V, in that space defined by six coordinates based on the attitudinal value of the six cognitive appraisals. V can be defined as:

$$V_A = \{ P, EF, C, AC, R, CO \} \qquad (4)$$

where, P, the dimension of pleasantness can be determined by:

$$P = \sum_{i=1}^{n} AP_{e_i} \qquad (5)$$

where $AP$ is the attitude of pleasantness towards element e in an event space consisting of $n$ elements. $EF$, $C$, $AC$, $R$, $CO$ are the dimensions of effort, certainty, activity, responsibility and control respectively and can be defined in a similar manner to *P*.

Having previously defined the event space in terms of its elements, we can determine the point in the affective space created by the six dimensional assessment of each element. The emotional reaction to the event is then calculated by finding the average of these points. This mechanism has been implemented in our emotionally motivated agent architecture or EMA.

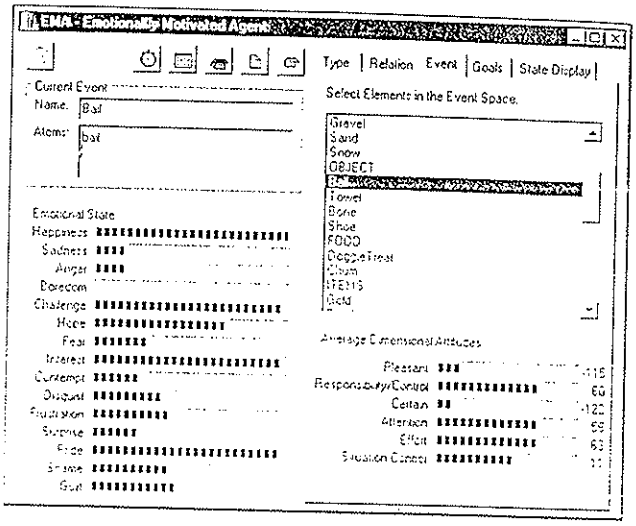
## EMA

EMA is driven by the goal hierarchy mechanism developed in its predecessor, GOMASE (Goal Orientated Multi-Agent Simulation Environment) (Baillie (Ray) and Lukose 1999). In the hierarchy, a goal can be either abstract or primitive. An abstract goal can be broken down into sub-goals (of which some will be abstract goals, while others may be primitive goals). Primitive or atomic goals correspond to an activity (or action) that needs to be carried out to achieve the goal. When a goal becomes the focus of an agent's belief and the agent wants to satisfy that goal, each sub-goal of that goal becomes active.

EMA may have any number of atomic goals for which it can perform tasks in order to satisfy a goal. In many cases where the agent has been given numerous task sets, not all of these need to be executed to satisfy the goal (Baillie (Ray) and Lukose 1999). Often when a subset of these tasks has been successfully completed the goal will be satisfied and the remaining active atomic goals and tasks can be deactivated.

Before making a decision, EMA generates an event space for each choice that it can make. From this event space, an emotional state is evoked based on the attitudinal values of the elements in the event space and where this places the agent within its affective space. Given the emotional states generated by each choice the agent makes an emotional decision about its choice(s). EMA makes a selection based on the most pleasant choice. This can be set to make the agent gravitate toward any emotion, preferred cognitive state or any combination of these.

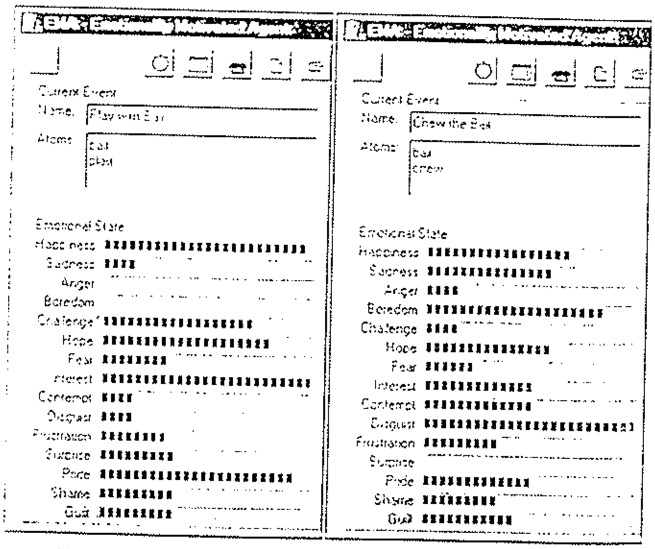
As the agent may have more than one goal active at any time, each goal is given a priority rating that is used by the agent to determine the urgency of an action. When there exist a number of solutions for a particular goal, the agent uses its emotional decision making abilities to then make a choice. To demonstrate this, we programmed EMA with the simple goal hierarchy of a dog (see Figure 5). Initially all EMA's goals are activated. One of which is the *play* goal. The activity that EMA evaluates to be the most pleasant and will satisfy the *play* goal is *dig hole*. Provided EMA has no other pressing goals, it will perform the action *dig hole*. As the simulation runs a built in energy gauge triggers the agents *eat* goal, simulating hunger. As time goes by and the *eat* goal is not satisfied it will become a higher priority than *play*. Although *dig hole* creates the most pleasant emotional state within the agent, it does not satisfy the highest priority goal of *eat*, therefore the agent selects the most pleasant action that it has in its goal hierarchy for satisfying the *eat* goal. To generate the emotional state of an activity the agent averages the attitudinal values of each element in the event space. This gives the agent a six dimensional point in its affective space. The distance of this point from all of the pure emotion points is calculated. Figure 3 shows an example of the agent's attitudinal values and generated emotional state for the concept *ball*. The emotional state generated by the concept *ball* is quite pleasant.[1]



Name: Ball
Atoms: ball

Pleasant -115
Responsibility/Control 60
Certain -122
Attention 59
Effort 63

Figure 3: Emotional State Evoked by Concept Ball



Name: Play with Ball
Atoms: ball
play

Name: Chew the Ball
Atoms: ball
chew

Figure 4: Emotional States of Play with Ball and Chew the Ball

However, an event does not only consist of objects. If the agent had to decide between *play ball* and *chew ball*, it would have to generate two event spaces, one for each action, and calculate the emotional state of each. Figure 4 show the agent's emotional states with respect to each of these actions.

Clearly the *play ball* action (shown on the left of Figure 4) produces a more pleasant emotional state than *chew ball*. In

---

[1] In keeping with Smith and Ellsworth's original model, the axis's of pleasantness and certainty are reversed, i.e. the lower the value the more pleasant or certain the concept is.

this case, the agent would select the *play ball* action. However, should the agent fail in its attempts to *play with ball* (for example the absence of someone to throw it), the agent will select its next most pleasant option, which in this case is *chew ball*. Thus, the most pleasant event available does not always produce a pleasant emotional state.

## INTERACTING WITH EMA

As previously mentioned, the agent was programmed with a simple dog-like goal hierarchy, illustrated in Figure 5. Programmed with appropriate trigger values for the goals (for example, *eat* is triggered by a threshold hunger value), the agent begins simulating behaviours applicable to achieving its goals.

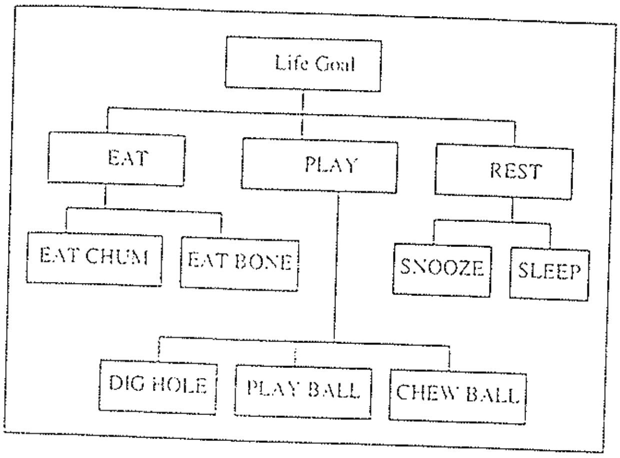


Figure 5: Goal Hierarchy for Dog-Like

The agent has no preconceived attitudes at the beginning of the simulation. As an action succeeds or fails in its attempts to achieve a goal, the agent's attitude values are updated and an emotional state is calculated by using the event spaces elementary attitudes to project a point into the affective space. The user can interact with the agent by placing different items in its environment. For example, the user can throw a ball for the agent to *play ball*. The user can also convey their attitudes about certain behaviours of the agent by *praising* (GOOD DOG) or *scorning* (BAD DOG) the agent during a particular behaviour. Table 1 is an extract of the narrative produced during the interactive simulation.

During the simulation the agent selects the activity based on the highest priority goal followed by its most pleasant choice. As can be seen in the narrative, the most pleasant choice at the time does not always produce pleasant emotions. In the beginning, all of the agent's goals are active. The agent wants to *eat* but cannot perform an appropriate action due to the absence of food in its environment. Next the agent tries to satisfy the *play* goal. Again, the environment is devoid of objects to play with. However the agent can satisfy the *play* goal by satisfactorily executing the *dig hole* action. Accordingly, because the *dig hole* action worked the first time, the agent selects this as it's most favourite action the next time it wants to satisfy the *play* goal. This action makes the agent *happy*. However the user scolds the agent after this activity, updating the agent's attitudes about this particul event space and thus changing the agent's emotional state.

```
Ema is looking for a bone.
Ema is looking for some chum.
Ema is hungry.
Ema wants to play.
Ema digs a hole.
She is feeling disgust
You dish up some chum.
Ema wants to play.
Ema digs a hole.
She is feeling happy
BAD DOG!
She is feeling contempt
Ema wants to play.
Ema is looking for a bone.
Ema is looking for some chum.
Ema finds some chum and eats it.
She is feeling contempt
You dish up some chum.
Ema wants to play.
Ema is tired.
Ema lies down for a rest.
She is feeling disgust
GOOD DOG!
She is feeling happy
Ema wants to play.
Ema digs a hole.
She is feeling disgust
BAD DOG!
She is feeling contempt
Ema wants to play.
Ema digs a hole.
She is feeling contempt
You throw the ball.
Ema wants to play.
Ema chews on the ball.
She is feeling contempt
BAD DOG!
She is feeling contempt
Ema wants to play.
Ema digs a hole.
She is feeling contempt
BAD DOG!

She is feeling contempt
Ema wants to play.
Ema fetches the ball.
She is feeling contempt
GOOD DOG!
She is feeling pride
Ema wants to play.
Ema chews on the ball.
She is feeling pride
BAD DOG!
She is feeling guilt
Ema is looking for some chum.
Ema finds some chum and eats it.
She is feeling guilt
GOOD DOG!
She is feeling happy
Ema wants to play.
Ema digs a hole.
She is feeling pride
BAD DOG!
She is feeling contempt
Ema wants to play.
Ema chews on the ball.
She is feeling contempt
BAD DOG!
She is feeling contempt
You throw the ball.
Ema wants to play.
Ema fetches the ball.
She is feeling contempt
GOOD DOG!
She is feeling pride
You throw the ball.
Ema wants to play.
Ema fetches the ball.
She is feeling happy
GOOD DOG!
She is feeling happy
You throw the ball.
Ema wants to play.
Ema fetches the ball.
She is feeling happy
```

Table 1: Narrative Output from EMA Simulation.

Although *dig hole* is now an unpleasant action for the agent, it is the only way that it can successfully achieve the *play* goal as the *ball* concept that is necessary for other actions is not in the agent's environment. Therefore, on successive *play* attempts, the agent selects the *dig hole* action. Eventually the user adds a *ball* to the environment. As the *ball* initially represents a neutral object in the agent's world, an activity that involves the *ball* is rated higher than the unpleasant *dig hole* action. On its first attempt the agent selects the *chew ball* action and is scolded. On the next attempt the agent selects *play ball*. The user praises the agent, the agent finds this a happy event. It attributes this happiness to the event space thus updating its attitudes towards the action *play* and the target object *ball*.

## DISCUSSION

The previous simulation is effective in demonstrating how the evaluation of an event space can produce a reasonable emotional state in the agent. In turn this emotional state is used to update the attitudinal values of elements in the event space. The agent has an internal mechanism that is used to trigger goals. It is a temporally sensitive energy gauge. It decreases steadily with time, but can be topped up or drained by certain activities. For example, eating or sleeping will increase the energy value but playing will decrease it. One threshold value on the gauge triggers the *eat* goal and another triggers the *sleep* goal. The *play* goal is intermittently triggered. When a goal is triggered, activities relating to this goal are placed in the agent's schedule where it proceeds to prioritize and sort the actions based on evoked emotions. The priority of a goal is determined by the number of times that goal has been triggered. For example, the *eat* goal is triggered by a threshold value on the energy gauge. As long as the gauge stays below that level, on each time increment in the simulation the *eat* goal will continue to be activated. When the agent satisfies the *eat* goal to a value less than the threshold value, the priority of *eat* is reset to zero and the goal is no longer invoked.

Each time an event occurs, the agent updates its values toward the six appraisal dimensions (i.e. *pleasantness, anticipated effort, certainty, attentional activity, responsibility* and *control*). Pleasantness is increased by a factor of the goal priority if the activity is completed successfully and increased if it fails. Effort is an evaluation of the rate of change in the energy gauge. Certainty is a calculation of the chance of success for a particular activity. Attention is calculated from the combined values of pleasantness, effort and goal priority. Responsibility is directly proportional to the goal priority and control is related to the success or failure of the activity.

The agent can use its emotional mechanisms to make adaptive decisions about not only situations that it is familiar with but also new situations that it has not encountered before. This is a desirable attribute in an agent.

## CONCLUSIONS AND FURTHER WORK

We believe that this atomic quantitative dimensional approach to synthesizing emotions in artificial beings is a powerful concept. In a world where many decision-making situations have numerous outcomes with an equal quality of solution, this type of reasoning is imperative in humans. Goleman in (Goleman 1995) states that emotional intelligence is the '*capacity to form an accurate, veridical model of oneself and to be able to use that model to operate effectively in life.*' If we expect to produce artificial intelligences with this human-like mentality, we need to explore the psychological theories of emotion, implementing them in artificial agent architectures and testing them for human-like affective decision-making capabilities.

In our EMA simulation, we have demonstrated the integration of a model proposed by Smith and Ellsworth (Smith and Ellsworth 1985) with our existing agent model. For preliminary experimentation we have chosen to simulate a simple pet dog as most people can identify with and have expectations on how a dog will behave and the behaviours and goals of a dog are much simpler than a person's are.

Our initial testing with this agent has been promising. Whilst it is still not clear how the six appraisal dimensions of the model should be manipulated during a simulation, our attempts show promise and do produce reasonable emotional responses. EMA still has a long way to go. We still need to address issues such as complex emotional states, how an emotional state affects a decision, how an emotional state should affect the agents perception and beliefs and the building of elaborate event spaces. To these ends, our research will continue its efforts on the EMA architecture and advance the understanding of emotions in the field of affective computing.

## REFERENCES

Baillie (Ray), P. and D. Lukose (1999). "To Believe, to Desire, to Intent: Motivating Autonomous Intelligent Agents with a Mind of their Own.". Managing Enterprises 99, Newcastle, School of Management, The University of Newcastle 235-240.

Baillie (Ray), P., D. Lukose, et al. (2000). "How to Give an Agent an Attitude". Agent Workshop 2000, Passau, Germany, Society for Computer Simulation International 7-12.

Baillie (Ray), P., M. Toleman, et al. (2000). "Could Emotions be the Key to Real Artificial Intelligence". ISA'2000, Wollongong, ICSC Academic Press 45-50.

Baillie, P., M. Toleman, et al. (2000). "Emotional Intelligence for Intuitive Agents". AISAT 2000, Hobart, University of Tasmania 134-139.

Damasio, A. R. (1994). "Descartes' Error: Emotion, Reason and the Human Brain". New York, N.Y., Gosset/Putnam Press.

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

El-Nasr, M. S. (1998). "Modelling Emotion Dynamics in Intelligent Agents". Cairo, American University in Cairo.

Fishbein, M. and I. Ajzen (1975). "Belief, Attitude, Intention and Behaviour". London, Addison-Wesley.

Goleman, D. (1995). "Emotional Intelligence". New York, Bantam Books.

Hanley, R. (1997). "The Metaphysics of Star Trek.". New York, Harper Collins Publishers, Inc.

Koestler, A. (1967). "The Ghost in the Machine". London, Penguin Books Ltd.

Petty, R. E. and J. T. Cacioppo (1996). "Attitudes and Persuasion: Classic and Contemporary Approaches". Boulder, Colorado, Westview Press.

Picard, R. (1998). "Affective Computing". London, The MIT Press.

Reilly, W. S. N. (1996). "Believable Social and Emotional Agents". School of Computer Science. Pittsburgh, Carnegie Mellon University: 288.

Smith, C. A. and P. C. Ellsworth (1985). "Attitudes and Social Cognition." Journal of Personality and Social Psychology 48(4): 813-838.

Smith, C. A. and P. C. Ellsworth (1985). "Patterns of Cognitive Appraisal in Emotion." Journal of Personality and Social Psychology 48(4): 813-838.