International Conference on Artificial Intelligence in Science and Technology
AISAT'2000

Hobart, Tasmania, Australia
17-20 December, 2000

EMOTIONAL INTELLIGENCE FOR INTUITIVE AGENTS

Penny Baillie*, Mark Toleman*, Dickson Lukose**

*Department of Mathematics and Computing, University of Southern Qld, Toowoomba, Qld 4350, AUSTRALIA. (ray@usq.edu.au | toleman@usq.edu.au)

**Brightware Inc., 90 Park Avenue, Suite 1600, New York, NY, 10016. USA. (dickson.lukose@brightware.com)

Abstract

Currently, there are no machines with emotions that influence their reasoning, perception and decision-making abilities to the degree that emotions affect human behaviour in these areas. This could be for two reasons. Firstly, emotions have traditionally been broadly defined and no discrete categorization had been formulated, and secondly, emotions have been viewed as opposing logic, the very basis for computational machines, and as a disruption to rational reasoning and function. It is the very contrasting evidence in recent research that has seen a renewed enthusiasm into emotional research. The role of emotion in rational human behaviour may have a larger impact on cognitive processes than first thought. In this paper, we define emotions and discuss the importance that they will have on artificial intelligences of the future.

1. INTRODUCTION

Dawkins, in [6] wrote, "A duck is a robot vehicle for the propagation of duck genes." This sentence could be construed somewhat controversial when viewed in context for AI researchers. When he wrote this, Dawkins was commenting on evolutionary and survival instincts in line with Charles Darwin's view that emotions being prevalent in both humans and other animals are crucial to the survival of the species [17].

If AI researches were to build an artificial robot duck, the least use they would have in mind for it would be the parenting of baby robot ducks. This broaches a fundamental question in the AI domain. Do emotions have any significant contributions to make to the area?

AI purists would agree with the traditional philosophical view of emotions, that they are an interruption to an otherwise logical state of being. However Hanley [11] distinguishes between the logical and emotional states and disagrees saying that without emotions we would be frozen into an eternal state of non-action due to the almost infinite number of logical actions that we need to choose from daily.

Picard, in [19], agrees with this view and sites a mass of current scientific evidence that emotions play an important part in cognitive processes [22] and that they impact significantly on rational thinking, perception, learning and other psychological functions that mold sentient behaviour. Her suggestion of an intuitive reasoning artificial intelligence to deal with "situations where problems cannot be enumerated and evaluated in the available time", is one of the motivations behind our work.

In recent years, a small group of dedicated researchers have developed artificial agents with affective reasoning and decision-making abilities based on appraisals of the agent's goals [7], [20]. We believe that while this method has proved effective, it generalizes an agent's emotional reactions and does not lend itself to emergent adaptive behaviour within the agent.

For this reason, our research has concentrated on the emotional assessment of atomic elements within the virtual world of an artificial agent. This approach is used to allow an agent to calculate an emotional attitude towards items within its world (see [3]).

To generate this type of behaviour in our agents we have integrated the psychological *theory of reasoned action* [18] into our agent architecture. This model examines the relationships of attitude between four key elements that a person interacts with when working towards attaining a goal. These elements are the action being performed, the target or targets that are the object of the action, the context of the action and, the temporal alignment of the action.

Our research relies heavily on the premise of Picard's *intuitive agent* and envisages a meshing of several psychological theories on emotion generation with agent technology, in order to produce an *intuitively* rational reasoning and decision making artificial being.

To this end, this paper is organized in the following manner: in Section 2, we will define emotion and look at two distinct categories; in Section 3, we discuss a relationship between attitude and appraisal theory for relationships between event appraisal and emotion generation; in Section 4, we briefly discuss our experimental emotional agent; and we conclude in Section 5, by summarizing our motivations in research and future directions.

2. DEFINING EMOTIONS

Emotion presents itself as a somewhat non-concise term in many of the domains that boast an understanding of the topic. These range from neurology [8], [17] and psychology [22] to artificial intelligence [16]. The reason may be that the term is used to describe a large range of cognitive and physiological states in sentient beings.

Emotions are often referred to in the broad sense, to describe not only familiar feelings such as happiness and sadness, but also biological motivational urges such as hunger and thirst [17]. Koestler, in [13], summed up this general view defining emotions as "mental states accompanied by intense feelings and involving bodily changes of a widespread character."

The degree of difficulty plaguing this subject matter is the fact that we very rarely experience a *pure* emotion. For example, feelings of hunger may be accompanied by feelings of frustration. However, there is a logical intuitive difference in defining hunger as an emotion and frustration as an emotion. Ask someone to contrast them and they will inevitably give you a look of, "well that is obvious", and begin, "hunger is a feeling..", followed by a perplexed look when they realize that that is the very way they are about to describe frustration.

In an attempt to isolate the mechanisms of emotion and devise a firmer definition we shall refer to two categories that logically divide emotions.

2.1 Survival Emotions

The first category is that of emotional drives. These drives or urges can be classified according to their source, pleasure rating and strength. This category of emotion has been particularly favourable with Freudian and Behaviourist Schools of psychology [10], [12]. The four emergent classifications for emotional drive can be identified as: emergency, biological, cognitive, and social.

MacLean [14] refers to the two basic emotional drives in animal behaviour; *self-preservation* and *preservation of the species*. From these two ultimate urges all motivational goals of animal behaviour can be derived and are individualistic of the being that they are being defined for [2].

An artificial agent system for modelling the effects of this category of emotion on behaviour is given in [1]. This system is discussed further in section 4 where we have run experiments using the survival emotions as motivation mechanisms to trigger an agent's goals. Here emotional drives are modelled as dynamic gauges with changing values depending on the state of the agent. When a threshold value for an emotional gauge is reached appropriate behaviours are triggered in the agent.

One example of a robotic device displaying this type of emotion is that of W. Grey Walter's *tortoise* mentioned in [16] which was capable of autonomous motion and had the ability to plug itself into a power socket to recharge its batteries. Although a simple decision making algorithm was obviously deployed to cause this behaviour, it nonetheless displayed *emergent emotions* and *emotional behaviour* [19].

This set of emotions we will refer to as *survival* or *lower* emotions (*drives* or *urges*). What may be inferred from the *tortoise* example above is that survival emotions are *emergent* and *inherent* of the type of survival needs of the individual being or device. For example, robotic devices will fail to operate without a continuing power source, therefore, the need for continued power and an awareness of the current power reserve state can be paralleled with human hunger. Without it, the robot would cease to function.

The second types of emotion that we will address are those of higher, neurological emotions.

2.2 Higher Emotions

If survival emotions are those that generate goals and motivation, then higher emotions are the resultant mental (and in turn physical) states generated by attempts to satisfy these goals. These emotions include *feelings* such as *happiness, anger, sorrow, guilt* and *boredom*.

Attempts to satisfy goals driven by the survival emotions give rise to an appropriate set of higher emotions. However, the emotions generated are difficult to categorize due to their complex relationships. Some emotions are viewed as being opposite such as *happiness* and *sadness*. These are easily distinguished because one is pleasant and the other unpleasant. On the other hand, *fear* and *anger* could also be considered opposite, but clearly not in the same way [21].

A popular theory for categorizing these emotions in psychology is that of cognitive appraisal theories [20], [19]. This theory lends itself ideally to application in computer science and AI due to the discrete nature of the categories.

Several models of appraisal for higher emotions have been developed. One that has been widely adopted for emotion synthesis in artificial beings is the OCC model [15]. This model assesses the human affective reaction to goal attainment as being either, pleasing, displeasing, approving, disproving, liking or disliking. The OCC cognitive appraisal model, and subsets of it, has been successfully implemented in a number of emotional agent architectures such as Elliott's Affective Reasoner [7] and Reilly's Believable and Emotional Agents [20].

Because the OCC model categorizes emotions based on goals as a whole it is difficult to address adaptive emotional behaviour stemming from attitudes about items that may interact with the individual during goal attainment and use this information to predict and extrapolate future behaviours under over lapping circumstances.

The need to identify the underlying elementary triggers that produce different emotional responses in different situations causes us to choose a deeper cognitive appraisal model. We consider Smith and Ellsworth's six dimensional cognitive appraisal model of emotions [21] to be more suitable for this application.

This dimensional theory identifies six orthogonal planes by which to determine an emotional category, these being, pleasantness, responsibility, certainty, attention, effort and situational control (see [21] for explanation). Within this 6 dimensional space, Smith and Ellsworth have identified the approximate empirical location of 15 separate higher emotions: *happiness, sadness, anger, boredom, challenge, hope, fear, interest, contempt, disgust, frustration, surprise, pride, shame and guilt.*

What our research focuses on are the attitudes toward the elementary items involved in goal attainment attempts. Our approach differs from other implementations of the cognitive appraisal theory in that it takes a micro perspective of an event where goal attainment is being attempted and evaluates it based on beliefs and the measurements of attitude that we apply to the elements involved in the event or what we call the *event space*.

3. ATTITUDE THEORY AND EVENT SPACES

While there is no clear definition of attitude, there is popular consensus among social psychologists that the term refers to the general enduring disposition to feel positively or negatively towards an object, person or issue [18]. By using a relevant collection of a person's attitudes, it is possible to closely predict the behaviour of that person where the same set of attitudes are applied [9]. This is otherwise known as *the theory of reasoned action*.

We can express an attitude towards an object, A as:

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

where W_i is the weighting of blame or praise associated with the object being instrumental in the failure or success of the goal i. V_i is the degree of satisfaction or dissatisfaction obtained from performing the goal and n is the number of times that the goal has been performed using the object.

Given a set of these elementary objects involved in an attempt to attain a goal, we define, what we have termed, the *event space* [3].

The event space E, can be defined as:

$$E = \{a, o, c, t\} \qquad 2$$

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.

In the following section we will briefly look at how attitude theory and the event space have been implemented in our agents and some results from experimentation.

With the results from the attitude appraisal of an event, we can synthesize appropriate higher level emotions by generating empirical values for each of the six appraisal dimensions identified in [21]. Each of these value combines to give us a six dimensional point located within the space, thus identifying an approximate emotional response to an event.

In turn, we need to examine these higher level generated emotions and identify how they should affect the agent's perception of information along with reasoning and decision-making abilities.

In the next section we look at how the theory of reasoned action and the event space have been integrated into our agent architecture to produce intuitive reasoning and decision-making mechanisms.

4. DEVELOPMENT OF AN EMOTIONAL AGENT

In our initial testbed environment GOMASE (Goal-Orientated, Multi-Agent Simulation Environment), agents are driven by a goal hierarchy [1] where individual goals are triggered by *lower emotional urges*. In this we assume that a goal can be either abstract or primitive. An abstract goal can be decomposed 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 beliefs and the agent wants to satisfy that goal, each sub-goal of that goal becomes active.

A GOMASE agent is also capable of developing attitudes about individual items within an event space. Each item is assessed on its contributions to the success or failure of goal attainment [2]. These attitudes are developed through the agent's experiences within its environment and are used to assess future events and make emotional decisions about future choices.

This gives the agent the ability to make choices in an environment where there is no correct answer. The agent uses attitudes from past experiences to select an option that it *likes* best. It calculates its total attitude about a choice based on its individual attitudes about the elements that make up the event space and prioritizes them in order on a scale from *liked* to least *disliked* with varying degrees in-between.

One experiment that we conducted was to give the agent one simple goal and several ways methods for satisfying this goal. The agent was preprogrammed with the knowledge that there were 10 different ways in which to satisfy this goal. Each solution had an equal degree of success or failure. If the agent chose a solution that failed it would attribute blame to the elements of that choice, if the agent succeeded its attitude about the elements would increase. The agent was run to satisfy this goal 100 times, each time retaining attitudes developed in past attempts. Data was gathered about the choices that the agent made.

This same goal was also given to a human subject. At this point it may be simpler to insert an extract of the narrative given to the human subject at the outset of the experiment in order to explain it.

> You are in a plane crash and find yourself stranded in the Jungle. It is 9am on a clear sunny morning. You have to find your way back to civilization. In the back of the plane you find 10 survival packs. Each survival pack contains a number of items to help you with your journey. Each item in the pack can be responsible for your success or failure. If just one item fails when you try to use it then your mission fails.

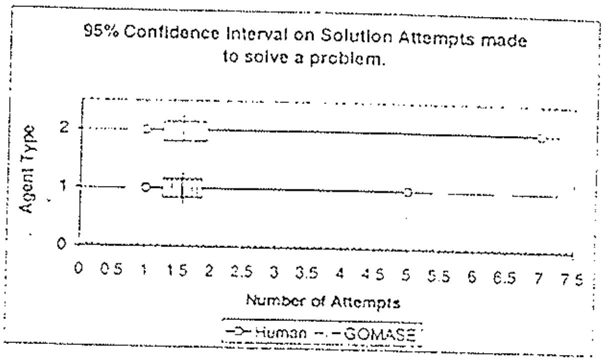

Figure 1. 95% C.I. on number of solutions attempts made before a successful option is found for the attainment of the same goal made by the GOMASE agent and a human agent

Figure 1 illustrates the GOMASE agent's ability to attain a goal given a large number of initially equal optimal solutions in a similar time frame to a human agent making the same decision.

On analysis of the final results, the GOMASE agent was as efficient at finding a solution as the human agent was. Details recorded about successful and unsuccessful attempts showed that without ever attempting all possible solutions, each agent was able to find and formulate attitudes about different solutions, thus *liking* them and preferring to choose that solution in a future attempt.

Although simple attitude theory is a successful vehicle to start our agents on the journey to affective reasoning and decision making, we still need to explore an extensive range of emotions and how the fundamental attitude appraisals of *liking* and *disliking* transform into broader and more complex emotions.

5. EMA

EMA, short for *emotionally motivated agent*, is our agent architecture with affective reasoning and decision-making abilities. EMA is an extension of the agent architecture designed and implemented in GOMASE. It integrates Smith and Ellsworth 6 dimensional model of emotion as discussed in Section 2 of this paper to produce an artificial intelligence capable of synthesizing *higher level emotions*.

This model [21], identifies 6 orthogonal dimensions: *pleasantness, anticipated effort, certainty, attentional activity, responsibility* and *control*, across 15 emotions: *happiness, sadness, anger, boredom, challenge, hope, fear, interest, contempt, disgust, frustration, surprise, pride, shame* and *guilt*.

Figure 2 displays two of the orthogonal dimensions and where the 15 emotional states are positioned with respect to them.

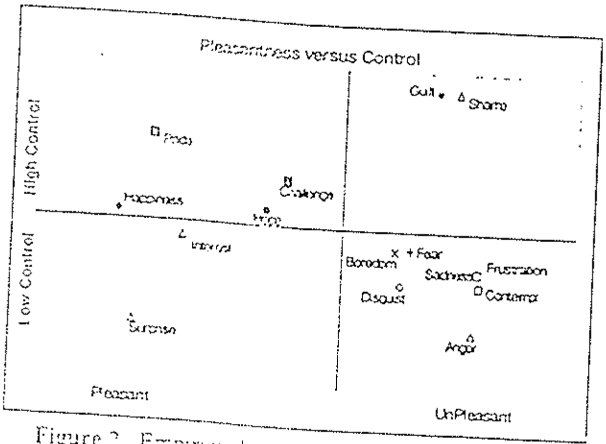

Figure 2. Empirical location of emotional states with respect to the pleasantness and control dimensions.

This research demonstrates a strong relationship between how human subjects interpret a past event and their own emotional reactions to it. It also provides us with an *affective* space, bounded by these dimensions, in which to explore the realm of emotion synthesis.

Like the GOMASE agent, EMA may have any number of atomic goals for which it can perform tasks in order to satisfy the goal. In many cases where the agent has been given numerous task sets, not all of these need to be executed to satisfy the [3]. 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.

Figure 3 displays the current user interface of EMA with a given situation.

In its present state EMA evaluates its emotional state based on a given event, the elements in that event and its attitude towards these elements. The given event/scenario is *taking Deefa the dog for a walk*.

The agent initially assesses its attitude towards each atomic element involved in the event. In this case the scenario includes the object *Deefa the dog* and the action *walk*. The user can set the agent's attitude values towards these elements in each of the 6 appraisal dimensions. This can be seen on the slide bars on the right-hand side of Figure 2.

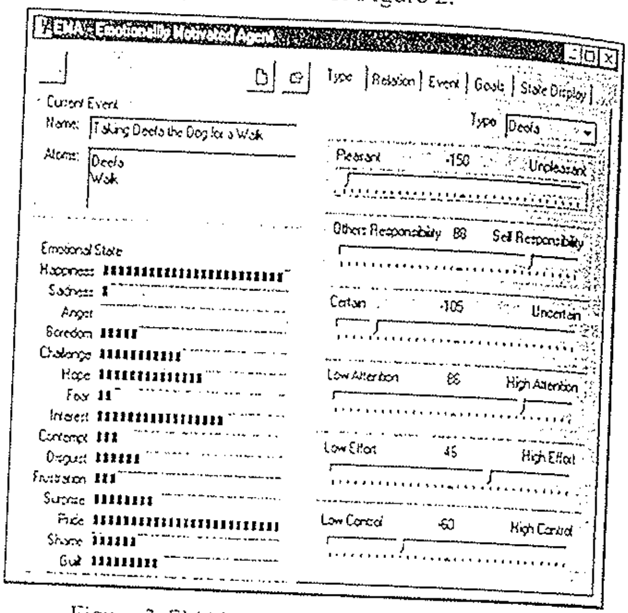

Figure 3 EMA's User Interface with Emotional

The total attitudes are combined using a simple averaging model [5]. This creates a single 6 dimensional coordinate in our affective space that is used to determine the artificial emotional state of the agent by calculating the distance of this point from each of the 15 emotion points.

6. CONCLUSIONS

In this paper we have discussed emotions and how they should be considered for integration into humanistic intelligent agents. We began our exploration of such an agent in the goal orientated, multi-agent simulation environment called GOMASE.

We are still researching the elements of attitude and event appraisal and how they relate to motivation in human beings using our Emotionally Motivated Agent. Currently, we are testing our agent's abilities at reasoning about situations and comparing them with human subject's reactions in the same virtual environment. This will help us to fine-tune our model and further develop the mechanism in our agents by which they appraise situations.

It may prove to be difficult to separate emotions into the two logical divisions of higher and lower affects, however if we are going to discard the flesh for the mind and instill emotional mechanisms into artificial beings, the distinction needs to be made. Therefore, we are not so concerned with the physiological and biological machines that process sensory input and generate emotional reactions, as we are with identifying algorithms that can synthesize emotional behaviour from similar sensory inputs and make it useful and believable.

Higher level emotions are a *fuzzy* area in both AI and physiological research [4]. We are yet to explore how these emotions will affect the agent's information processing capabilities and how the outward behaviour of the agent will be perceived. We do know, however, that these emotions should affect and distort the agent's perception of the world in which it is exists [4].

To these ends, our continuing research and the development of the E.M.A. software will endevour to expand our understanding of attitude theory, cognitive appraisal theory and its influence on affective reasoning in humans and the translation of these models into an affective agent architecture.

7. REFERENCES

[1] Baillie (Ray), P. and D. Lukose. To Believe, to Desire, to Intent: Motivating Autonomous Intelligent Agents with a Mind of their Own. in Managing Enterprises 99. 1999. Newcastle.

[2] Baillie (Ray), P. and D. Lukose, Urging Desire: Motivational Mechanisms for Intelligent Agents with Minds of Their Own. Cybernetics and Systems, 2000. Special Edition.

[3] Baillie (Ray), P., D. Lukose, and M. Toleman. *How to Give an Agent an Attitude*. in *Agent Workshop 2000*. 2000. Passau, Germany.

[4] Carley, K. and A. Newell, *The Nature of the Social Agent*. Journal of Mathematical Sociology, 1994. 19(4): p. 221 - 262.

[5] Davidson, A.R., *From Attitudes to Actions to Attitude Change: The Effects of Amount and Accuracy of Information*, in *Attitude Strength*, R.E. Petty and J.A. Krosnick, Editors. 1995, Lawrence Erlbaum Associates: Mahwah, New Jersey.

[6] Dawkins, R., *The Selfish Gene*. 1976, Oxford: Oxford University Press.

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

[8] Fellous, J., *The Neuromodulatory Basis of Emotion*. The Neuroscientist, 1999. Summer 1999.

[9] Fishbein, M. and I. Ajzen, *Belief, Attitude, Intention and Behaviour*. 1975, London: Addison-Wesley.

[10] Furth, H., *Knowledge As Desire, An Essay on Freud and Piaget*. 1987, New York: Columbia University Press.

[11] Hanley, R., *The Metaphysics of Star Trek*. 1997, New York: Harper Collins Publishers, Inc.

[12] Hull, C.L., *Principals of behaviour*. 1943, New York: Appleton-Century-Crofts.

[13] Koestler, A., *The Ghost in the Machine*. 1967, London: Penguin Books Ltd.

[14] MacLean, P., Contrasting Functions of Limbic and Neocortical Systems of the Brain and their Relevance to the Psycho-physical Aspects of Medicine. American Journal of Medicine, 1958. XXV(4): p. 6711 - 6726.

[15] Ortony, A., G.L. Clore, and A. Collins, *The Cognitive Structure of Emotions*. 1988, Cambridge, UK.: Cambridge University Press.

[16] Penrose, R., *The Emperor's New Mind*. 1989, New York: Oxford University Press.

[17] Pert, C.B., *Molecules of Emotion*. 1997, New York: Simon and Schuster.

[18] Petty, R.E. and J.T. Cacioppo, *Attitudes and Persuasion: Classic and Contemporary Approaches*. 1996, Boulder, Colorado: Westview Press.

[19] Picard, R., *Affective Computing*. 1998, London: The MIT Press.

[20] Reilly, W.S.N., *Believable Social and Emotional Agents*, in *School of Computer Science*. 1996, Pittsburgh: Carnegie Mellon University.

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

[22] Smith, C.A. and L.D. Kirby, eds. *Consequences Require Antecedents: Towards a Process Model of Emotion Elicitation*. Feeling and Thinking: The role of affect in social cognition., ed. J. Forgas. 2000, Cambridge University Press: London.