FLAME - Fuzzy Logic Adaptive Model of Emotions

Authors : Magy Seif El-Nasr (contact person), John Yen, Thomas R. Ioerger

Contact Information:

Snail Mail:

Computer Science Department

Texas A&M University

College Station, TX 77844-3112

Emails:

magys@cs.tamu.edu

ioerger@cs.tamu.edu

yen@cs.tamu.edu

Tel:

409-862-9243

Fax:

409-862-9243

Abstract

Emotions are an important aspect of human intelligence and have been shown to play a significant role in the human decision-making process. Researchers in areas such as cognitive science, philosophy, and artificial intelligence have proposed a variety of models of emotions. Most of the previous models focus on an agent's reactive behavior, for which they often generate emotions according to static rules or pre-determined domain knowledge. However, throughout the history of research on emotions, memory and experience have been emphasized to have a major influence on the emotional process. In this paper, we propose a new computational model of emotions that can be incorporated into intelligent agents and other complex, interactive programs. The model uses a fuzzy-logic representation to map events and observations to emotional states. The model also includes several inductive learning algorithms for learning patterns of events, associations among objects, and expectations. We demonstrate empirically through a computer simulation of a pet that the adaptive components of the model are crucial to users' assessments of the believability of the agent's interactions.

1 Introduction

Emotions, such as anger, fear, relief, and joy, have long been recognized to be an important aspect of the human mind. However, the role that emotions play in our thinking and actions has often been misunderstood. Historically, a dichotomy has been perceived between emotion and reason. Ancient philosophers did not regard emotion as a part of human intelligence, but rather they viewed emotion as an impediment — a process that hinders human thought. Plato, for example, said "passions and desires and fears make it impossible for us to think" (in *Phaedo*). Descartes echoed this idea by defining emotions as passions or needs that the body imposes on the mind, and suggesting that they keep the mind from pursuing its intellectual process.

More recently, psychologists have begun to explore the role of emotions as a positive component in human cognition and intelligence (Bower and Cohen 1982, Ekman 1992, Izard 1977 and Konev et al. 1987). A wide variety of evidence has shown that emotions have a major impact on memory, thinking, and judgment (Bower and Cohen 1982, Konev et al. 1987, Forgas 1994 and Forgas 1995). For example, neurological studies by Damasio and others have demonstrated that people who lack the capability of emotional response often make poor decisions that can seriously limit their functioning in society (Damasio 1994). Gardner proposed the concept of "multiple intelligences." He described personal intelligence as a specific type of human intelligence that deals with social interaction and emotions (Gardner 1983). Later, Goleman coined the phrase "emotional intelligence" in recognition of the current view that emotions are actually an important part of human intelligence (Goleman 1995).

Many psychological models have been proposed to describe the emotional process. Some models focus on the effect of motivational states, such as pain or hunger. For example, Bolles and Fanslow proposed a model to account for the effect of pain on fear and vice versa (Bolles and Fanslow 1980). Other models focus on the process by which events trigger certain emotions; these models are called "event appraisal" models. For example, Roseman et. al. developed a model to describe emotions in terms of distinct event categories, taking into account the certainty of the occurrence and the causes of an event (Roseman et al. 1990). Other models examine the influence of expectations on emotions (Price et al. 1985). While none of these models presents a complete view, taken as a whole, they suggest that emotions are mental states that are selected on the basis of a mapping that includes a variety of environmental conditions (e.g., events) and internal conditions (e.g., expectations, motivational states).

Inspired by the psychological models of emotions, Intelligent Agents researchers have begun to recognize the utility of computational models of emotions for improving complex, interactive programs. For example, interface agents with a model of emotions can form a better understanding of the user's moods, emotions and preferences and can thus adapt itself to the user's needs (Elliot 1992, Maes 1997). Software agents may use emotions to facilitate the social interactions and communications between groups of agents (Dautenhahn. 1998), and thus help in coordination of tasks, such as among cooperating robots (Shibata et al. 1996). Synthetic characters can use a model of emotion to simulate and express emotional responses, which can effectively enhance their believability (Bates 1992a, Bates 1992b, and Bates et al. 1992). Furthermore, emotions can be used to simulate personality traits in believable agents (Rousseau 1996).

One limitation that is common among the existing models is the lack of adaptability. Most of the computational models of emotions were designed to respond in pre-determined ways to specific situations. The dynamic behavior of an agent over a sequence of events is only apparent from the change in responses to situations over time. A great deal of psychological evidence points to the importance of memory and experience in the emotional process (LeDoux 1996 and Ortony et al. 1988). For example, classical conditioning was recognized by many studies to have major effects on emotions (Bolles and Fanselow 1980, and LeDoux 1996). Consider a needle that is presented repeatedly to a human subject. The first time the needle is introduced, it inflicts some pain on the subject. The next time the needle is introduced to the subject, he/she will typically expect some pain, hence he/she will experience fear. The expectation and the resultant emotion are experienced due to the conditioned response. Some psychological models explicitly use expectations to determine the emotional state, such as Ortony et al.'s model (Ortony et al. 1988). Nevertheless, classical conditioning is not the only type of learning that can induce or trigger expectations. There are several other types of learning that need to be incorporated to produce a believable adaptive behavior, including learning about sequences of events and about other agents or users.

In this paper, we propose a new computational model of emotions called FLAME, for "Fuzzy Logic Adaptive Model of Emotions." FLAME is based on several previous models, particularly those of Ortony et al. (Ortony et al. 1988) and Roseman et al.'s (Roseman et al. 1990) event-appraisal models, and Bolles and Fanselow's (Bolles and

Fanslow 1980) inhibition model. However, there are two novel aspects of our model. First, we use fuzzy logic to represent emotions by intensity, and to map events and expectations to emotional states and behaviors. While these mappings can be represented in other formalisms, such as the interval-based approach used by the OZ project (Reilly 1996), we found that fuzzy logic allowed us to achieve smooth transitions in the resultant behavior with a relatively small set of rules. Second, we incorporate machine learning methods for learning a variety of things about the environment, such as associations among objects, sequences of events, and expectations about the user. This allows the agent to adapt its responses dynamically, which will in turn increase its believability.

To evaluate the capabilities of our model, we implemented a simulation of a pet named PETEEI - a PET with Evolving Emotional Intelligence. We performed an ablation experiment in which we asked users to perform tasks with several variations of the simulation, and then we surveyed their assessments of various aspects of PETEEI's behavior. We found that the adaptive component of the model was critical to the believability of the agent within the simulation. We argue that such a learning component would also be equally important in computational models of human emotions, though they would need to be extended to account for interactions with other aspects of intelligence. We then address some limitations of the model and discuss some directions for future research on FLAME.

2 Previous Work

Models of emotion have been proposed in a broad range of fields. In order to review these models, we have grouped them according to their focus, including those emphasizing motivational states, those based on event appraisals, and those based on computer simulations. In the next few sections, we will discuss examples of each of these types of models.

2.1 Motivational States

Motivational states are any internal states that promote or drive the subject to take a specific action. In this paper, we consider hunger, fatigue, thirst, and pain as motivational states. These states tend to interrupt the brain to call for an important need or action (Bolles and Fanslow 1980). For example, if a subject is very hungry, then his/her brain will direct its cognitive resources to search for food, which will satisfy the hunger. Thus, these states have a major impact on the mind, including the emotional process and the decision-making process, and hence behavior.

Models of motivational states, including pain, hunger, and thirst, were explored in various areas of psychology and neurology (Schumacher and Velden 1984). Most of these models tend to formulate the motivational states as a pure physiological reaction, and hence the impact that these motivational states have on other processes, such as emotions, has not been well established. A model was proposed by Bolles and Fanselow (1980) to explore the relationship between motivational states and emotional states, specifically between "fear" and "pain." Their idea was that motivational states sometimes inhibit or enhance emotional states. For example, a wounded rat that is trying to escape from a predator is probably in a state of both fear and pain. In the first stage, fear inhibits pain to allow the rat to escape from its predator. This phenomenon is caused by some hormones that are released when the subject is in the fear state (Bolles and Fanslow 1980). At a later stage, when the cause of the fear disappears (i.e., the rat successfully escapes) and the fear level decays, pain will inhibit fear (Bolles and Fanslow 1980), hence causing the rat to tend to its wounds. In some situations, pain was found to inhibit fear and in others fear was found to inhibit pain. The model emphasized the role of inhibition and how the brain could suppress or enhance some motivational states or emotions over others. This idea was incorporated as part of our model, but covers only one aspect of the emotional process.

2.2 Appraisal Models and Expectations

As another approach to understanding the emotional process, some psychologists tried to formulate emotions in terms of responses to events. The models that evolved out of this area were called "event appraisal" models of emotions. Roseman et al. (1990) proposed a model that generates emotions according to an event assessment procedure. They divided events into motive-consistent and motive-inconsistent events. Motive-consistent events are events that are consistent with one of the subject's goals. On the other hand, a motive-inconsistent event refers to an event that threatens one of the subject's goals. Events were further categorized with respect to other properties. For example, an event can be caused by other, self or circumstance. In addition to the knowledge of the cause, they assessed the certainty of an event based on the expectation that the event would actually occur. Another dimension that was used to differentiate some emotions was whether an event is motivated by the desire to obtain a reward or avoid a punishment. To illustrate the importance of this dimension, we use "relief" as an example. They defined relief as an emotion that is triggered by the occurrence of a motive-consistent event which is motivated by avoiding punishment. In other words, relief can be defined as the occurrence of an event that avoids punishment. Another important factor that was emphasized was self-perception. The fact that subjects might regard themselves as weak in some situations and strong in some others may trigger different emotions. For example, if an agent expected a motive-inconsistent event to occur with a certainty of 80% and it regarded itself as weak in the face of this event, then it will feel fear. However, if the same situation occurred and the agent regarded itself as strong, then frustration will be triggered (Roseman et al. 1990).

This model, like most of the event appraisal models of emotion, does not provide a complete picture of the emotional process. The model does not describe a method by which perceived events are categorized. Estimating the probability of occurrence of certain events still represents a big challenge. Furthermore, some events are perceived contradictorily as both motive-consistent and motive-inconsistent. In this case, the model will produce conflicting emotions. Therefore, a filtering mechanism by which emotions are inhibited or strengthened is needed. Moreover, the emotional process is rather deeply interwoven with the reasoning process, among other aspects of intelligence, and thus, not only external events, but also internal states trigger emotions.

Ortony et al. (1988) developed another event-appraisal model that was similar to Roseman's model, but used a more refined notion of goals. They divided goals into three types: A-goals defined as preconditions to a higher-level goal; I-goals defined as implicit goals such as life preservation, well being, etc.; and R-goals defined as explicit short-term goals such as attaining food, sleep, water, etc. They also defined some global and local variables that can potentially affect the process by which an emotion is triggered. Local variables were defined to be: the likelihood of an event to occur, effort to achieve some goal, realization of a goal, desirability for others, liking of others, expectations, and familiarity. Global variables were defined as sense of reality, arousal, and unexpectedness. They used these terms to formalize emotions. For example: joy = the occurrence of a desirable event, relief = occurrence of a disconfirmed undesirable event. Sixteen emotions were expressed in this form, including relief, distress, disappointment, love, hate and satisfaction (Ortony et al. 1988).

Nevertheless, this model, as Roseman's model, does not provide a complete picture of the emotional process. As stated earlier the rules are intuitive and seem to capture the process of triggering individual emotions well, but often emotions are triggered in a mixture. The model does not show how to filter the mixture of emotions triggered to obtain a coherent emotional state. Since the model was developed for understanding emotions rather than simulating emotions, the calculation of the internal local and global variables, such as the expectation or likelihood of event occurrence, was not described.

Although Roseman et al.'s (1990) and Ortony et al.'s (1988) models demonstrated the importance of expectations, they did not identify a specific link between expectations and the intensity of the emotions triggered. To quantify this relationship, D. Price and J. Barrell (1985) developed an explicit model that determines emotional intensities based on desires and expectations. They asked subjects questions about their experiences with various emotions, including anger and depression. They then developed a mathematical curve that fit the data collected. They generalized their findings into a quantitative relationship among expectation, desire and emotional intensity. However, the model did not provide a method for acquiring expectations and desires. Still, it confirmed the importance of expectation in determining emotional responses, and we were able to use some of their equations in our model.

2.3 Models of Emotions in AI

Through the history of Artificial Intelligence (AI) research, many models have been proposed to describe the human mind. Several models have been proposed to account for the emotional process. Simon developed one of the earliest models of emotions in AI (Simon 1967). Essentially, his model was based on motivational states, such as hunger and thirst. He simulated the process in terms of interrupts; thus whenever, the hunger level, for example, reaches a certain limit, the thought process will be interrupted. R. Pfeifer (1988) has summarized AI models of emotions from the early 1960's through the 1980's. However, since the psychological picture of emotions was not well developed at that time, it was difficult to build a computational model that captures the complete emotional process. In more recent developments, models of emotions have been proposed and used in various applications. For example, a number of models have been developed to simulate emotions in decision-making or robot communication (Sugano and Ogata 1996, Shibata et al. 1996, Breazeal and Scassellati to appear, Brooks et al. to appear). In the following paragraphs, we will describe applications of models of emotions to various AI fields, including Intelligent Agents.

Bates' OZ project

J. Bates built believable agents for the OZ project (Reilly and Bates 1992, Bates et al. 1992a and Bates et al. 1992b) using Ortony et al.'s event-appraisal model (Ortony et al. 1988). The aim of the OZ project was to provide users with the experience of living in dramatically interesting micro-worlds that include moderately competent emotional agents. They formalized emotions into types or clusters, where emotions within a cluster share similar causes. For example, the distress type describes all emotions caused by displeasing events. The assessment of the displeasingness of events is based on the agent's goals. They also mapped emotions to certain actions.

The model was divided into three major components: TOK, HAP and EM. TOK is the highest-level agent within which there are two modules: HAP is a planner and EM models the emotional process. HAP takes emotions and attitudes from the EM model as inputs and chooses a specific plan from its repository of plans, and carries it out in the environment. It also passes back to the EM component some information about its actions, such as information about goal failures or successes. EM produces the emotional state of the agent according to different factors, including the goal success/failure, attitudes, standards and events. The emotional state is determined according to the rules given by Ortony et al.'s (1988) model. Sometimes the emotions in EM override the plan in HAP, and vice versa. The outcome of the selected plan is then fed back to the EM model, which will reevaluate its emotional status. Thus, in essence, emotions were used as preconditions of plans (Reilly and Bates 1992).

Many interesting aspects of emotions were addressed in this project; however the underlying model still has some limitations. Even though it employed Ortony's emotional synthesis process (Ortony et al. 1988), which emphasized the importance of expectation values, the model did not attempt to simulate the dynamic nature of expectations. Expectations were generated statically according to predefined rules. Realistically, however, expectations change over time. For example, a person may expect to pass a computer science course. However, after taking few computer science courses and failing them, his/her expectation of passing another computer science course will be much lower. Therefore, it is very important to allow expectations to change with experience. In our model, which we discuss in the next section, we incorporated a method by which the agent can adapt its expectations according to past experiences.

Cathexis Model

A model, called Cathexis, was proposed by Velasquez (1997) to simulate emotions using a multi-agent architecture. The model only described basic emotions and innate reactions, however it presented a good starting point for simulating emotional responses. Some of the emotions simulated were anger, fear, distress/sadness, enjoyment/happiness, disgust, and surprise. The model captures several aspects of the emotional process, including (1) neurophysiology, which involves neurotransmitters, brain temperatures, etc., (2) sensorimotor aspect, which models facial expressions, body gestures, postures and muscle action potentials, (3) a simulation of motivational states and emotional states, and (4) event appraisals, interpretation of events, comparisons, attributions, beliefs, desires, and memory. The appraisal model was based on Roseman et al.'s (1990) model. The model handles mixtures of emotions by having the more intense emotions dominate other contradictory ones. In addition, emotions were decayed over time. The model did not account for the influence of motivational states, such as pain (Bolles and Fanslow 1980). Moreover, the model did not incorporate adaptation in modeling emotions. To overcome these limitations, we used several machine learning algorithms in our model and incorporated a filtering mechanism that captures the relations between emotions and motivational states.

Elliot's Affective Reasoner

Another multi-agent model, called Affective Reasoner, was developed by C. Elliot (Elliot 1992, Elliot 1994). The model is a computational adaptation of Ortony et al.'s psychological model (Ortony et al. 1988). Agents in the Affective Reasoner project are capable of producing twenty-four different emotions, including joy, happy-for, gloating, resentment, sorry-for, and can generate about 1200 different emotional expressions. Each agent included a representation of the self (agent's identity) and the other (identity of other agents involved in the situation). During the simulation, agents judge events according to their pleasantness and their status (unconfirmed, confirmed, disconfirmed). Joy, for example, is triggered if a confirmed desirable event occurs. Additionally, agents take into account other agents' responsibility for the occurring events. For example, gratitude towards another agent can be triggered if the agent's goal was achieved, i.e. a pleasant event, and the other agent is responsible for this achievement. In addition to the emotion generation and action selection phases, the model presents another dimension to emotional modeling, which is social interaction. During the simulation, agents, using their own knowledge of emotions and actions, can infer the other agents' emotional states from the situation, from their emotional expressions, and their actions. These inferences can potentially enhance the interaction process (Elliot 1992).

Even though Elliot's model presents an interesting simulation describing emotion generation, emotional expressions, and their use in interactions, the model still faces some difficulties. The model does not address several issues, including conflicting emotion resolution, the impact of learning on emotions or expectations, filtering emotions, and their relation to motivational states. Our model, described below, addresses these difficulties.

Blumberg's Silas

Another model which is related to our work, is Blumberg's (1996) model. Blumberg is developing believable agents that model life-like synthetic characters. These agents were designed to simulate different internal states, including emotions and personality (Blumberg 1996). Even though he developed a learning model of both instrumental and classical conditioning, which were discussed in (Dojman 1998), he did not link these learning algorithms back to the emotional process generation and expression (Blumberg et al. 1996). His work was directed toward using learning as an action-selection method. Even though we are using similar types of learning algorithms (e.g. reinforcement learning), our research focuses more on the impact of this and other types of learning on emotional states directly.

Rousseau's CyberCafe

To model a believable agent for interacting with humans in a virtual environment, one will have to eventually consider personality as another component that has to be added to the architecture of the model. Rousseau has developed a model of personality traits that includes, but is not limited to, introverted, extroverted, open, sensitive, realistic, selfish, and hostile (Rousseau 1996). The personality traits were described in terms of inclination and focus. For example, an open character will be greatly inclined to reveal details about him/herself (i.e., high inclination in the revealing process), while an honest character will focus on truthful events when engaging in a revealing process (i.e., high focus on truth). An absent minded character will be mildly inclined to pay attention to events (i.e., low inclination in the perception process), while a realistic character focuses on the real or confirmed events. Additionally, they looked at the influence of personality on several other processes including moods and behavior (Rousseau 1997, Rousseau and Barbara Hayes-Roth 1996). Moods were simulated as affective states that include happiness, anger, fatigue, and hunger, which are a combination of emotional and motivational states in our model. Our treatment and definition of mood, which will be discussed later, is quite different.

3. Proposed Model

3.1 Overview of the Model's Architecture

In this section, we describe the details of a new model of emotions called FLAME - Fuzzy Logic Adaptive Model of Emotions. The model consists of three major components: an emotional component, a learning component and a decision-making component. Figure 1 shows an abstract view of the agent's architecture. As the figure shows on the right-hand side, the agent first perceives external events in the environment. These perceptions are then passed to both the emotional component and the learning component (on the left-hand side). The emotional component will process the perceptions; in addition, it will use some of the outcomes of the learning component, including expectations and event-goal associations, to produce an emotional behavior. The behavior is then returned back to the decision-making component to choose an action. The decision is made according to the situation, the agent's mood, the emotional states and the emotional behavior; an action is then triggered accordingly. We do not give a detailed model of the action-selection process, since there are a number of planning or rational decision-making algorithms that could be used (Russell and Norvig 1995). In the following sections, we describe the emotional component and learning component in more detail.

3.2. Emotional Component

3.2.1. Overview Of The Emotional Process

The emotional component is shown in more detail in Figure 2. In this figure, boxes represent different processes within the model. Information is passed from one process to the other as shown in the figure. The perceptions from the environment are first evaluated. The evaluation process consists of two sequential steps. First, the experience model determines which goals are affected by the event and the degree of impact that the event holds on these goals. Second, mapping rules compute a desirability level of the event according to the impact calculated by the first step and the importance of the goals involved. The event evaluation process depends on two major criteria: the importance of the goals affected by the event, and the degree by which the event affects these goals. Fuzzy rules are used to determine the desirability of an event according to these two criteria.

The desirability measure, once calculated, is passed to an appraisal process to determine the change in the emotional state of the agent. FLAME uses a combination of Ortony et al.'s (1988) and Roseman et al.'s (1990) models to trigger emotions. An emotion (or a mixture of emotions) will be triggered using the event desirability measure. The mixture will then be filtered to produce a coherent emotional state. The filtering process used in FLAME is based on Bolles and Fanslow (1980)'s approach, described in more detail below. The emotional state is then passed to the behavior selection process. A behavior is chosen according to the situation assessment, mood of the agent, and the emotional state. The behavior selection process is modeled using fuzzy implication rules. The emotional state is eventually decayed and fed back to the system for the next iteration. Additionally, there are other paths by which a behavior can be produced. Some events or objects may trigger a conditioned behavior, and thus these events might not pass through the normal paths of the emotional component (LeDoux 1996).

3.2.2. Use of Fuzzy Logic

Motivated by the observation that human beings often need to deal with concepts that do not have well-defined sharp boundaries, Lotfi A. Zadeh developed fuzzy set theory that generalizes classical set theory to allow the notion of partial membership (Zadeh 1965). The degree an object belongs to a fuzzy set, which is a real number between 0 and 1, is called the membership value in the set. The meaning of a fuzzy set is thus characterized by a membership function that maps elements of a universe of discourse to their corresponding membership values.

Based on fuzzy set theory, fuzzy logic generalizes modus ponens in classical logic to allow a conclusion to be drawn from a fuzzy if-then rule when the rule's antecedent is partially satisfied. The antecedent of a fuzzy rule is usually a boolean combination of fuzzy propositions in the form of ``x is A" where A is a fuzzy set. The strength of the conclusion is calculated based on the degree to which the antecedent is satisfied. A fuzzy rule-based model uses a set of fuzzy if-then rules to capture the relationship between the model's inputs and its output. During fuzzy inference, all fuzzy rules in a model are fired and combined to obtain a fuzzy conclusion for each output variable. Each fuzzy conclusion is then defuzzified, resulting in a final crisp output. An overview of fuzzy logic and its formal foundations can be found in (Yen 1999).

FLAME uses fuzzy sets to represent emotions, and fuzzy rules to represent mappings from events to emotions, and from emotions to behaviors. Fuzzy logic provides an expressive language for working with both quantitative and qualitative (i.e., linguistic) descriptions of the model, and enables our model to produce some complex emotional states and behaviors. For example, the model is capable of handling goals of intermediate importance, and the partial impact of various events on multiple goals. Additionally, the model can manage problems of conflicts in mixtures of emotions (Elliot 1992). Though these problems can be addressed using other approaches, such as functional or interval-based mappings (Velasquez 1997, Reilly 1996), we chose fuzzy logic as a formalism mainly due the simplicity and the ease of understanding linguistic rules. We will describe below the fuzzy logic models used in FLAME.

3.2.3 Event Evaluation

We use fuzzy rules to infer the desirability of events from its impact on goals, and the importance of these goals. The impact of an event on a goal is described using five fuzzy sets: *HighlyPositive, SlightlyPositive, NoImpact SlightlyNegative* and *HighlyNegative* (see Figure 3). The importance of a goal is dynamically set according to the agent's assessment of a particular situation. The importance measure of a goal is represented by three fuzzy sets: *NotImportant, SlightlyImportant* and *ExtremelyImportant* (see Figure 4). Finally, the desirability measure of events can be described as *HighlyUndesired, SlightlyUndesired, Neutral, SlightlyDesired,* and *HighlyDesired* (see Figure 5).

To determine the desirability of events based on their impact on goals and the goals' importance, we used fuzzy rules of the form given below:

IF Impact(G_1,E) is A_1

AND Impact(G_2,E) is A_2

…..

AND Impact(G_k,E) is A_k

AND Importance(G_1) is B_1

AND Importance(G_2) is B_2

….

AND Importance(G_k) is B_k

THEN Desirability(E) is C

where *k* is the number of goals involved. A_i, B_j, and *C* are represented as fuzzy sets, as described above. This rule reads as follows: if the goal, G_1, is affected by event *E* to the extent A_1 and goal, G_2, is affect by event *E* to the extent A_2, etc., and the importance of the goal, G_1, is B_1 and the importance of goal, G_2, is B_2, etc., then the desirability of event *E* will be *C*.

We will use an example to illustrate how these fuzzy rules are used in our model. Consider an agent personifying a pet. An event, such as taking the food dish away from the pet may affect several immediate goals. For example, if the pet was hungry and was planning to reach for the food dish, then there will be a negative impact on the pet's goal to prevent starvation. It is thus clear that the event (i.e., taking away the dish) is undesirable in this situation. The degree of the event's undesirability is inferred from the impact of the event on the starvation prevention goal and the importance of the goal. Thus, the rule relevant to this situation is:

IF Impact(prevent starvation, food dish taken away) is *HighlyNegative*

AND Importance(prevent starvation) is *ExtremelyImportant*

THEN Desirability(food dish taken away) is *HighlyUndesired*

There are several different types of fuzzy rule-based models: (1) the Mamdani model (Mandani and S. Assilian 1975), (2) the Takagi-Sugeno model (Takagi and Sugeno 1985), and (3) Kosko's Standard Additive Model (Kosko 1997). We chose the Mamdani model with centroid defuzzification. The Mamdani model uses *Sup-Min* composition to compute the matching degrees for each rule. For example, consider the following set of *n* rules:

If x is A_1 Then y is C_1

…..

If x is A_n Then y is C_n

where x is an input variable, y is an output variable, A_i and C_i are fuzzy sets, and i represents the ith rule. Assuming, the input x is a fuzzy set A', represented by a membership function $\mu_{A'}(x)$ (e.g. degree of impact). A special case of A' is a singleton, which represents a crisp (non-fuzzy) input value. Given that, the matching degree w_i between the input $\mu_{A'}(x)$ and the rule antecedent $\mu_{A_i}(x)$ is calculated using the equation below:

$$\sup_X(\mu_{A_i}(x) \wedge \mu_{A'}(x))$$

The $\wedge$ operator takes the minimum of the membership functions and then a *sup* operator is applied to get the maximum over all x. The matching degree affects the inference result of each rule as follows:

$$\mu_{C'_i}(y) = w_i \wedge \mu_{C_i}(y)$$

where C'_i is the value of variable y inferred by the *i*th. fuzzy rule. The inference results of all fuzzy rules in the Mamdani model are then combined using the max operator $\vee$ (i.e., the fuzzy disjunction operator in the Mamdani model):

$$\mu_{comb}(y) = \mu_{C'_1}(y) \vee \mu_{C'_2}(y) \vee \ldots \vee \mu_{C'_k}(y)$$

This combined fuzzy conclusion is then defuzzified using the following formula based on center of area (COA) defuzzification:

$$y_{final} = \frac{\int \mu_{comb}(y) \quad y \quad dy}{\int \mu_{comb}(y) \quad dy}$$

The defuzzification process will return a number that will then be used as a measure of the input event's desirability.

3.2.4 Event Appraisals

Once the event desirability is determined, rules are fired to determine the emotional state, which also takes into account expectations. Expectation values are derived from the learning model, which is detailed in a later section. Relationships between emotions, expectations, and the desirability of an event are based on the definitions presented by Ortony et al. (1988), are given in Table 1. Fourteen emotions were modeled. Other emotions like love or hate towards another are measured according to the actions of the other and how it helps the agent to achieve its goals. To implement the rules shown in the table, we need the following elements:

- the desirability of the event, which is taken from the event evaluation process discussed in the previous section,
- standards and event judgment, which are taken from the learning process, and
- expectations of events to occur, which are also taken from the learning process.

To illustrate the process, we will employ the emotion of relief as an example. Relief is defined as the occurrence of an disconfirmed undesirable event, i.e. the agent expected some event to occur and this event was judged to be undesirable, but it did not occur. The agent is likely to have been in a state of fear in the previous time step, because fear is defined as expecting an undesirable event to happen. A history of emotions and perceived events are kept in what is called the short-term emotional memory. Thus, once a confirmed event occurs, it is checked with the short-term emotional memory; if a match occurs, a relative emotion is triggered. For example, if the emotion in the previous time step was fear, and the event did not occur, then relief will be triggered. The intensity of relief is then measured as a function of the prior degree of fear.

The quantitative intensities of emotions triggered by these rules can be calculated using the equations formulated by Price et al. (1985). For example, hope is defined as the occurrence of an unconfirmed desirable event, i.e. the agent is expecting a desirable event with a specific probability. Consider a student who is repeating a course is expecting to get an A grade with a probability of 80%. The hope intensity is not directly proportional to the expectation value, as might be the case with other emotions. On the contrary, the higher the certainty, the less the hope (Price et al. 1985). The hope intensity can be approximated by:

$$Hope = (1.7 \times expectation^{0.5}) + (-0.7 \times desirability)$$

Other formulas for emotions in our model are shown in Table 2. The table shows the method by which intensities are calculated for various emotions given an expectation value and an event desirability measure.

Emotions such as pride, shame, reproach and admiration, which do not depend directly on expectations and desirability, are functions the agent's standards. The calculation of the intensity of any of these emotions will depend primarily on the value of the event according to the agent's acquired standards. For example, if the agent learned that a given action, *x*, is a good action with a particular goodness value, *v,* then if the agent causes this action, *x*, in the future, it will experience the emotion of pride, with a degree of *v*.

3.2.5 Emotional Filtering

Emotions usually occur in mixtures. For example, the feeling of sadness is often mixed with shame, anger or fear. Emotions are sometimes inhibited or enhanced by other states, such as motivational states. Bolles and Fanslow's (1980) work gives an insight on the impact that motivational states may have on other emotions. In their work the highest intensity state dominates (Bolles and Fanslow 1980). With emotional states this might not necessarily be the case. For example, a certain mixture of emotions may produce unique actions or behaviors. In the following paragraph, we will provide a description of how we simulate the interaction between the motivational states and the emotional states. We note that, in general, emotional filtering may be domain-dependent and influenced by other complicated factors such as personality.

Our method of filtering emotions relies on motivational states. Motivational states tend to interrupt the cognitive process to satisfy a higher goal. In our simulation of a pet described in Section 4, these states include, hunger, thirst, pain, and fatigue. Table 3 shows the different motivational states of the pet that are simulated, and the factors determining their intensities. These motivational states have different fuzzy sets representing their intensity level, e.g. *LowIntensity*, *MediumIntensity* and *HighIntensity*. Once these states reach a sufficient level, say *MediumIntensity*, they send a signal to the cognitive process indicating a specific need that has developed. These motivational states can then block the processing of the emotional component to produce a plan to enable the agent to satisfy its needs, whether it is water, food, sleep, etc. The plan depends on the agent's situation at the particular time. For example, if the agent already has access to water, then it will drink, but if it does not have access to water and knows its whereabouts, then it will form a plan to get it. However, if the agent must depend on another agent to satisfy its needs then it will use the model that it learned about the other agent to try to manipulate it. It is not always best for the agent to inhibit emotional states to achieve a goal or satisfy a need. Sometimes the agent will be acting on fear and inhibiting other motivational states. The emotional process always looks for the best emotion to express in various situations. In some situations, it may be best if fear inhibits pain, but in some others it may not (Bolles and Fanslow 1980). According to Bolles and Fanselow's model, fear inhibits pain if (1) the cause of fear is present and (2) the fear level is higher than the pain level. At a later time step, when the cause of fear disappears, pain will inhibit the fear. Thus, before inhibiting emotions, we make a situation assessment and an emotional versus motivational states assessment. Whichever is best for the agent to act on in the given situation will take precedence, while the others will be inhibited.

Inhibition can occur directly between emotions as well, i.e. sadness or anger may inhibit joy or pride in some situations. Some models employ techniques that tend to suppress weaker opposite emotions (Velasquez 1997). For example, an emotion like sadness will tend to inhibit joy if sadness was more intense than joy. Likewise, if joy was more intense than anger or sadness, then joy will inhibit both. In our model, we employ a similar technique. Thus, if joy was high and sadness was low, then joy will inhibit sadness. However, we give a slight preference to negative emotions since they often dominate in situations where opposite emotions are triggered with nearly equal intensities.

Mood may also aid in filtering the mixture of emotions developed. Negative and positive emotions will tend to influence each other only when the mood is on the boundary between states (Bower and Cohen 1982). Moods have been modeled by others, such as in Cybercafe (Rouseau and Hayes-Roth 1997). However, in these models the mood was treated as a particular affective state, such as fatigue, hunger, happiness and distress. In contrast, our model simulates the mood as a modulating factor that can either be positive or negative. The mood depends on the relative intensity of positive and negative emotions over the last n time periods. (We used n=5 in our simulation, because it was able to capture a coherent mixture of emotional states). Opening up the time window for tracking moods may cause dilution by more conflicting emotions, while a very narrow window might not average over enough emotional states to make a consistent estimate. We calculate the mood as follows:

$$mood = \begin{Bmatrix} pos & if \sum_{i=-n}^{-1} I_i^+ > \sum_{i=-n}^{-1} I_i^- \\ neg & otherwise \end{Bmatrix}$$

Where I_i^+ is the intensity of positive emotions at time *i*, and I_i^- is the intensity of negative emotions at time *i*. To illustrate the calculation of the mood, we will employ the following example. In this example, the mood is negative, resulting from three negative emotions and two positive emotions all with a medium intensity. The mixture of the emotions triggered was as follows: (1) a positive emotion, *joy*, with a high intensity (0.25) and (2) a negative emotion, *anger*, with a relatively lower intensity (0.20). The negative emotion inhibits the positive emotion despite the fact that the positive emotion was triggered with a higher intensity, because the agent is in a negative mood. We use a tolerance of ±5% to define the closeness value of two intensities (through trial and error 5% has been shown to produce adequate results with the pet prototype). Thus, if one emotion has a value of, *l*, then any emotions with a value of *l±5%* will be considered close, and hence will depend on mood to be the deciding factor.

3.2.6 Behavior Selection

Fuzzy logic is used once again to determine a behavior based on a set of emotions. The behavior depends on the agent's emotional state and the situation or the event that occurred. For example, consider the following rule:

If Anger is High
AND dish-was-taken-away
THEN behavior is Bark-At-user

The behavior, *Bark-At-User*, depends on what the user did and the emotional intensity of the agent. If the user did not take the dish away and the agent was angry for some other reason, it would not necessarily be inclined to bark at the user, because the user might not be the cause of its anger. Thus, it is important to identify both the event and the emotion. It is equally important to identify the cause of the event. To generalize the rule shown above, we used the following fuzzy rules:

IF emotion$_1$ is A$_1$
AND emotion$_2$ is A$_2$
…..
AND emotion$_k$ is A$_k$
AND Event is E
AND Cause (E, B)
THEN BEHAVIOR is F

where k is the number of emotions involved. A_1, A_2 and A_k are fuzzy sets defining the emotional intensity as being *HighIntensity, LowIntensity or MediumIntensity*. The event is described by the variable E and the cause of the event is described by the variable B. Behaviors are represented as singletons (discrete states), including *Bark-At-User* and *Play-With-Ball*. Likewise, events are simulated as singletons such as *dish-was-taken-away, throw-ball, ball-was-taken-away*, etc. In the case of PETEEI, we are assuming that non-environmental events, such as *dish-was-taken-away*, *throw-ball*, *ball-was-taken-away*, etc. are all caused by the user. Using the fuzzy mapping scheme, the behavior with the maximum value will be selected.

To elaborate on how behaviors are selected in the model, we will present an example involving fear and anger. Consider, for instance, every time you take the food dish away from the dog you hit it to prevent it from jumping on you and barking at you after taking its food away. Taking the food dish away produces anger, because the pet will be experiencing both distress and reproach, and as shown in Table 1, anger is a compound emotion consisting of both reproach and distress. The pet will feel reproach because, by nature, it disapproves of the user's action (taking the food dish away), and it will be distressed because the event is unpleasant. Additionally, since the user hits the dog whenever he/she takes the dish away, fear will be produced as a consequence. Thus, taking the dish away will produce both anger and fear. Using fuzzy rules, the rule fired will be as follows:

IF Anger is *HighIntensity*
AND fear is *MeduimIntensity*
And Event is *dish-was-taken-away*
THEN BEHAVIOR is growl.

Therefore, the behavior was much less aggressive than with anger alone. In effect, the fear dampened the aggressive behavior that might have otherwise been produced.

3.2.7 Decay

At the end of each cycle (see figure 2), a feedback procedure will reduce the agent's emotions and reflect them back to the system. This process is important for a realistic emotional model. Normally, emotions do not disappear once their cause has disappeared, but rather they decay through time, as noted in (Velasquez 1997). However, very few studies have addressed the emotional decay process. In FLAME, a constant, ϕ, is used to decay positive emotions, and another constant, δ, is used to decay negative emotions. Emotions are decayed toward 0 by default. For example: $I_{e_i}(t+1) = \phi \bullet I_{e_i}(t)$ for positive emotions, e_i. We set $\phi < \delta$, to decay positive emotions at a faster rate, since intuitively negative emotions seem to be more persistent. This choice was validated by testing the different decay strategies using an agent-based simulation. These constants, along with the constants used in the learning algorithm, were passed as parameters to the model. We used trial and error to find the best settings for these parameters. We found that there was a range of settings that produced a reasonable behavior for the agent. These ranges were: $0.1 < \phi < 0.3$ and $0.4 < \delta < 0.5$.

3.3 Learning Component

3.3.1. Overview of the Learning Component

Learning and adaptability can have a major impact on emotional dynamics. For example, classical conditioning was shown to play a significant role in determining emotional responses (LeDoux 1996). Classical conditioning is not the only type of learning that can impact the emotional process. In fact, through our research we have discovered many other types of learning to be important for modeling the emotional intelligence process, including learning actions that please or displease the user and learning about what events to expect. Recognizing the importance of each of these types to the emotional process, we added a learning component to FLAME. To simulate the different learning types, we employed different inductive techniques, including (1) conditioning to associate an emotion with an object that had triggered the emotion in the past, (2) reinforcement learning to assess events according to the agent's goals, (3) a probabilistic approach to learn patterns of events, and (4) a heuristic approach to learn actions that please or displease the agent or the user.

3.3.2 Classical Conditioning

Associating objects with emotions or with a motivational state forms a simple type of learning in FLAME. For example, if the agent experiences pain when an object, *g*, touches it, then the motivational state of pain will be associated with the object *g*. This kind of learning does not depend on the situation per se, but it depends on the object-emotion/motivational state association.

Each of these associations will have an accumulator, which is incremented by the repetition and intensity of the object-emotion occurrence. This type of learning will provide the agent with a type of expectation triggered by the object, rather than the event. Using the count and the intensity of the emotion triggered, the agent can calculate the expected intensity of the emotion. We used the formula shown below:

$$\langle I(e) | o \rangle = \frac{\sum\limits_{events(i)} I_i(e)}{n_o}$$

where *events(i)* are events that involve the object *o*, $I_i(e)$ is the intensity of emotion *e* in event *i*, and n_o is the total number of events involving object *o*. In essence, the formula is averaging the intensity of the emotion in the events where the object, *o,* was introduced. To illustrate this process we give the following example. Consider a needle that is introduced to the agent. The first time the needle is introduced it causes the agent 30% pain. So the agent will associate the needle with 30% pain. The next time we introduce the needle, the agent will expect pain with a level of 30%. Let's say that we introduce the needle 99 more times without inflicting any pain. Then, the next time the needle is introduced, the agent will expect pain with a level of $\frac{0.3 \times 1}{100} = 0.3\%$. As you can see, the level of expectation of the intensity of a particular emotion, *e*, is decreased by the number of times the object, *o*, was introduced without inflicting the emotion, *e*. If, however, we shock the agent with the needle 90 times (30% of pain each time) and only introduce it 10 times without inflicting any pain, then the expectation of pain will be increased: $\frac{90 \times 0.3}{100} = 27\%$.

Using these associations, emotions can be triggered directly. In the example of the needle, when the needle is introduced, the emotion of fear will be automatically triggered because the agent will be expecting pain. However, if the object was associated with sadness or joy, an emotion of sadness or joy might be triggered in these cases, respectively. The intensity of these emotions will be calculated using the formula described above.

3.3.3 Learning about Impact of Events

In addition to direct associations formed by the classical conditioning paradigm, the agent needs to learn the general impact of events on its goals. It is often the case that a given event does not have an impact on any specific goal directly, but instead, some sequence of events may eventually have an impact on a goal. However, identifying the link between an event and the affected goals has been noted to be a very complex task (Reilly and Bates 1992), since the agent often does not know the consequences of a given action until a complete sequence of actions is finished. The agent, therefore, faces the problem of temporal credit assignment, which is defined as determining which of the actions in a sequence is responsible for producing the eventual rewards. The agent can potentially learn this by using reinforcement learning (Mitchell 1996). We will briefly outline a reinforcement algorithm, namely Q-learning. The reader is referred back to (Kaelbling et al. 1996) for more detail. It should be noted that Q-learning is just one approach to learn about events; other approaches may be used to equally get the desired effect. For example, Blumberg (Blumberg et al. 1996) has developed a model that uses temporal difference learning to develop a mechanism for classical and instrumental conditioning, which has been discussed by many psychologists, especially those in the neuroscience and the animal learning fields, such as Rescorla (1988 and 1991).

To illustrate the solution that reinforcement learning offers to this problem, we will look at Q-learning in more detail. The agent represents the problem space using a table of "Q-values" in which each entry corresponds to a state-action pair. The table can be initially filled with default values (for example, 0). The agent begins from an initial state *s*. It takes an action, *a*, by which it arrives to a new state s'. The agent may obtain a reward, *r*, for its action. As the agent explores its environment, it accumulates observations about various state transitions, along with occasional rewards. With each transition it updates the corresponding entry in the Q-table above using the following formula:

$$Q(s,a) \leftarrow r + \gamma \max_{a'} Q(s',a')$$

where *r* is the immediate reward, γ is a discount factor ($0 \leq \gamma < 1$), and the a' are the actions that can be taken from the new state s'. Thus, the Q-value of a state-action pair depends on the Q-values of the new state. After many iterations, the Q-values converge to values that represent the expected payoff in the long-run for taking a given action in a given state, which can be used to make optional decisions.

This basic algorithm is guaranteed to converge only for deterministic Markov Decision Processes (MDP)[1]. However, in our application there is an alternation between actions of the agent and the user, introducing non-determinism. The reward for a given state and action may change according to the user and environment. For example, at a state, s_0, suppose the agent does an action, a_0, and as a consequence the user gives him a positive reward. At a later time step, the agent is in the same state, s_0, so he takes action, a_0, again but this time the user decides to reward him negatively. Therefore, the user introduces a non-deterministic response to the agent's actions. In this case, we treat the reward as a probability distribution over outcomes based on the state-action pair. Figure 6 illustrates this process in more detail. The agent starts off in state, s_0, and takes an action a_0. Then the user can take either action, a_1, that puts the agent in state, s_1, or action, a_2, that puts the agent in state, s_2. The dotted lines show the nondeterminism induced by user's actions, while the straight black lines shows the state-action transition as represented in the table. The probability of the user's actions can be calculated using a count of the number of times the user did this action given the state-action pair, (s_0, a_0), divided by the number of actions the user took given that the agent was at state, s_0, and did action, a_0: $P(s'|s,a)$.

---

[1] A Markov Decision Process refers to an environment in which rewards and transition probabilities for each action in each state depend only on that state, and are independent of the previous actions and states used to get to the current state.

To calculate the Q-values given this nondeterministic model, we used the following formula:

$$Q_n(s,a) \leftarrow E[r(s,a)] + \gamma \sum_{s'} P(s'|s,a) \max_{a'} Q(s',a')$$

where $P(s'|s,a)$ is the probability that was described above, and $E[r(s,a)]$ is the expected reward (averaged over all previous executions of a in s). The summation represents the expected maximum Q-value over all possible subsequent states. We discuss learning the probabilities $P(s'|s,a)$ in the next subsection.

At any given time, the agent will be faced with different actions to take, each of which will result in different outcomes and different rewards. The formula and the algorithm described above gives the maximum expected reward, given that the agent is in a particular state, which is used to decide what the optimal action is to take. However, since we are trying to simulate a believable agent, we also want to account for other influences on how humans make decisions, such as the effect of moods (Bower and Cohen 1982, and Damasio 1994). We incorporated mood by modifying the expectation values of the next state, s', given that the agent is in state s. Instead of calculating the value of an action by maximizing the expected Q-value, we use the mood as a weighting factor to modify the expected probabilities of new states. As noted in (Bower and Cohen 1982), when the agent is in a positive mood it will tend to be more optimistic, so it will naturally expect desirable events to occur. To simulate this phenomenon, we altered the expectation formula described above. In a particular situation, the agent will be looking at numerous alternative actions, some of which may lead to failure and others may lead to success. If the agent's mood is positive then the agent will be optimistic and expect the good events to occur with a degree of β more than the bad events and vice versa. We modified the equation above to include this β value. Thus, the formula was revised as follows:

$$Q_n(s,a) \leftarrow E[r(s,a)] + \gamma \sum_{Match} (1+\beta) P(s'|s,a) \max_{a'} Q(s',a') + \gamma \sum_{nonMatch} \alpha P(s'|s,a) \max_{a'} Q(s',a')$$

Where *Match* denotes the states consistent with the particular mood, which was calculated as the average of the emotions in the last five time steps as described earlier. For example, if the mood is "good" then the states with a "good" assessment will fall under the *Match* category, while the states with a "bad" assessment will fall under the *nonMatch* category. We solve for α so that the weights still sum to 1:

$$\sum_{match} (1+\beta) P(s'|s,a) + \sum_{nonMatch} \alpha P(s'|s,a) = 1$$

$$\alpha = \frac{1}{\sum_{nonMatch} P(s'|s,a)} \left(1 - (1+\beta) \sum_{Match} P(s'|s,a)\right)$$

For all the actions that lead to a state that matches the mood, the value will be augmented by β. We will give an example to illustrate this idea. Suppose the agent is trying to decide between two actions. These actions are illustrated in Figure 7. If it ignores the ball then there is a 70% chance that it will end up in state, s_2, which has a max Q-value of 2.5, but there is also a 30% chance of ending up in state, s_1, which has a max Q-value of -1.5. While if it plays with the ball then there is an equal chance of getting to state, s_4, which has a max Q-value of -3, or state, s_5, which has a max Q-value of 5. Thus, if we use regular probability calculation, action *Ignore(Ball)* will have an expected Q-value of $0.3 \times -1.5 + 0.7 \times 2.5 = 1.3$, and *PlayWith(Ball)* will have a value of $0.5 \times -3 + 0.5 \times 5 = 1$. However, if we take the mood into account, then the calculation will be different because the agent will expect different outcomes with different moods. For example, if the agent is in a positive mood, it will expect positive rewards: *Give(Food)* and *Give(Drink)* with a degree β more than the negative rewards: *Take(Ball)* and *Talk("Bad Boy")*. If we set β to 50%, then *Ignore(Ball)* will have an expected Q-value of

$$0.3 \times -1.5 \times \alpha + 0.7 \times 2.5 \times (1+\beta) = 2.7$$

where $\beta = 0.5$, and $\alpha = \frac{1}{0.3}(1 - (1.5 \times 0.7)) = -0.167$

and *PlayWith(Ball)* will have a value of

$$0.5\times -3\times \alpha + 0.5\times 5\times (1+\beta) = 3\,.$$

where $\beta = 0.5$, and $\alpha = \frac{1}{0.5}(1-(1.5\times 0.5)) = 0.5$

Thus, the positive mood (β=50%) makes a slight change in the Q-values, causing *PlayWith(Ball)* to become more desirable than *Ignore(Ball)*, which could, in turn, affect the triggering of emotions or choice of actions by the agent.

3.3.4 Forming a User Model

The agent needs to know what events to expect, how likely they are to occur, and how bad or good they are. This information is crucial for the process of generating emotions. As was discussed in the sections above, the generation of emotions and emotional intensities relies heavily on expectations via event appraisals. While some researchers treat this as pre-determined knowledge built into the agent, we attempt to learn these expectations dynamically.

Since the agent is interacting with the user, the agent will have to learn about the user's patterns of behavior. A probabilistic approach is used to learn patterns based on the frequency with which an action, a_1, is observed to occur given that previous actions a_2, etc. have occurred. We focused on patterns of length three, i.e. three consecutive actions by the user. In the pet domain, actions are simple and quick, so sequences of one or two actions are often not meaningful. However, to be realistic, we did not want to require the pet simulation to maintain too many items in short-term memory (even humans appear to be limited to 7±2 items (Miller 1956)). So we restricted the learning of patterns to sequences of length three. A typical pattern is illustrated as follows: the owner goes into the kitchen (a_1), takes out the pet's food from the cupboard (a_2) and feeds the pet (a_3). These three consecutive actions led to the pet being fed. Thus, if the owner goes to the kitchen again, the pet would expect to be fed with some probability. In general, learning sequences of actions can be very useful in predicting a user's actions.

In FLAME we keep a table of counts which is used to define the conditional probability $p(e_3 | e_1, e_2)$ denoting the probability of an event e_3 to occur, given that events e_1 and e_2 have just occurred. When a pattern is first observed, an entry is created in the table that indicates the sequence of three events, with a count of 1. Then, every time this sequence is repeated, the count is incremented. We can use these counts to calculate the expected probability of a new event *Z* occurring, given that two previous events *X* and *Y* occurred. The expected probability of event *Z* is calculated as follows:

$$P(Z \mid X,Y) = \frac{C[X,Y,Z]}{\sum_i C[X,Y,i]}$$

Cases where experience is limited (e.g. number of relevant observations in low) can be handled by reducing the probability to be conditioned only on one previous event. For example, if the sequence *Y* and *Z* was observed, then the probability of event *Z* occurring is calculated as:

$$P(Z \mid Y) = \frac{\sum_i C[i,Y,Z]}{\sum_j \sum_i C[i,Y,j]}$$

However, if the prior event *Y* has never been seen to precede the event *Z*, the probability of event *Z* occurring can be calculated as an unconditioned prior (i.e., average probability over all events):

$$P(Z) = \frac{\sum_{i,j} C[i,j,Z]}{\sum_{i,j,k} C[i,j,k]}$$

These probabilities allow the agent to determine how likely the user is to take certain actions, given the events that have recently occurred, which is required for the reinforcement learning. It should be noted that the patterns discussed above only account for the user's actions, since the goal was to adapt to and learn about the tendencies of the user. A useful but more complex extension would be to interleave the agent's reactions in the patterns.

3.3.5 Learning Values of Actions

The learning algorithms described so far allow the agent to associate expectations and rewards to events. Although reinforcement learning can effectively be used to learn about rewards and expectations, which are used for making decisions to maximize expected payoff in the long-run, it is not sufficient to simulate emotions such as pride, shame or admiration. Such emotions seem to be based on a more immediate reflection of the goodness or badness of actions in the view of other agents or the user. Reinforcement learning, as described above, tends to produce a selfish agent (an agent looking at the desirability of events according to his own goals). Such a model is not sufficient to simulate the effects of emotions in social interaction. For example, it has been hypothesized by some social psychologists that in relationships, whenever partners feel guilty, they go into a submissive role, and they will try to make up for what they did wrong (Friske 1992). In this situation they will be using their emotional intelligence to search for an action that is pleasing to the partner (for example, giving a gift to the partner). Therefore, an agent that is engaging in social interaction with a user or another agent will have to use a similar protocol. However, to emulate this protocol, the agent will have to know what is pleasing and displeasing for other agents or for the user. Therefore, it would not only need to know how useful or desirable an action is to itself, but how desirable the action is to others. Additionally, it should be able to assess the values of its own actions to avoid hurting others.

To learn values of actions, we devised a simple new learning algorithm to associate the agent's actions with the user's feedback. We assume that the user can take some actions to provide feedback, such as saying "Bad Dog." The learning algorithm calculates an expected value for each of the agent's action using this user feedback. The algorithm averages feedback from the user on immediately preceeding actions, accumulated over a series of observations. For example, consider an agent who takes an action *BarkAt*(User). The user may as a consequence take actions, such as *YellAt*(agent) or *Hit*(agent). Each user action or feedback is assumed to have a value assigned in the agent's knowledge-base according to its impact; for example, *YellAt*(agent) might be worth -2 and *Hit*(agent) might be worth -6. Positive feedback are represented by positive numbers and negative feedback are represented by negative numbers. Using these values, the agent can track the average value of its own actions. The expected value of an action a is calculated as the sum of the values of user feedback given after each occurrence of action a over the number of occurrence of action a, which is formalized as follows:

$$value(a) = \frac{1}{|A|} \sum_{e \in A} feedback(e+1)$$

Where A represents set of events where the agent takes action a, and $e+1$ represents user's response in the next event. The agent can then use this formula to assess it's own actions and trigger evaluative emotions, such as pride or shame. Additionally, it can trigger emotions such as admiration or reproach by assessing the value of the other agent's actions according to its own standards. The agent could also use these expectations to determine actions that will be pleasing to the user or to other agents.

Figure 8 summarizes the modes of learning and their interactions with the emotional component in the FLAME model.

4. Simulation and Results

4.1 Simulation

In this section, we will describe an implementation of FLAME in an interactive simulation of a pet. Our simulation is called PETEEI — a PET with Evolving Emotional Intelligence. We chose to model emotions in a pet based on the following reasons:

- Pets are simpler than humans. They do not require sophisticated planning. Concepts of identity, self esteem, self-awareness and self-perception do not exist (or are not as pronounced) in animals. In addition, the goal structure of pets is much simpler than human's.
- A pet's behavior is relatively easy to evaluate. We thought it would be better to have a model of a pet rather than some other creature, because most people have expectations about what a pet should or should not do, and our evaluation will take advantage of this common knowledge.

PETEEI is implemented in Java with a graphical interface. It has five major scenes: a garden, a bedroom, a kitchen, a wardrobe and a living room. The garden scene is illustrated in Figure 9a.

In order to anticipate the various behaviors and situations that needed to be simulated, a list of user actions was defined, and is summarized as follows:

- *Walk to different scenes:* The user can walk from one scene to the next by clicking on the walk button and then clicking on the direction he wants to go to. The cursor will then change to an arrow with the destination name on it. For example, if the user clicked left and if the next scene to the left is the bedroom, then the cursor will be shaped as an arrow with the word bedroom on it.
- *Object Manipulation:* The user can take objects from a scene. This action will automatically add the object to the user's inventory, from which objects can be taken and introduced to other scenes.
- *Talk aloud:* The user can initiate a dialogue with objects (including the pet). Talk is done by selecting words from a predefined set of sentences, which are defined within the main window of the application. After selecting what the user wants to say, he/she can then click on talk. This will send a global message to all the objects within the scene, as if the user were talking aloud.
- *Opening and closing doors:* The user can open and close doors or other objects that may be opened or closed in the environment.
- *Look at:* The user can look at or examine various objects within the scene.
- *Touch and Hit:* The user can also touch or hit any object within the scene.

Feedback from the pet consists of barking, growling, sniffing, etc. (only sounds were used, no animation). In addition, there was a text window that described the pet's actions (looking, running, jumping, playing, etc.) , and a graphical display of internal emotional levels (shown in figure 9b). The graphical display of emotions was used in the evaluation process for two reasons. First, the pet's expressions or actions only represent emotional reactions; there was no planning or extended action involved in the model simulation. The model as it is produces only emotional behaviors, which is merely a subset of a more complex behavioral system. Therefore, asking the users to rate the pet's believability according to its actions only is insufficient to validate the hypothesis regarding emotions. Second, we are evaluating the simulation's emotional generation capability, rather than action selection. The focus of our model was to simulate emotional states, which could eventually be used in many different ways in different applications, such as synthetic character simulations, communication and negotiations in multi-agent systems, etc. Thus, evaluating the agent's actions does not suffice to validate the emotional model underneath the agent's simulation. In an attempt to validate both the emotional mappings and the action mappings, we present the user with both aspects and let him/her judge the model. For a complete list of questions, scenarios and introduction given to the user the reader is referred to (Seif El-Nasr 1998).

4.2 Evaluation Method

4.2.1. Evaluation Protocol and Validity

To evaluate the impact of various components of FLAME on PETEEI, we chose a method of user assessment in which users walked through different scenarios with the simulation, and then we gathered feedback via a questionnaire (Baecker et al. 1995). We gave users some predefined sequences of actions to perform within the simulation. When the users were finished, they were asked to answer a set of questions. The reason for using questionnaires as our evaluation method was three-fold. First, questionnaires provide users with structured answers to the questions that we are looking for. Second, questionnaires can be given to ordinary users, as opposed to soliciting more sophisticated feedback from experts (on pet behavior or social interaction). Finally, having users assess the behavior of a pet is more convenient than comparing the performance of the simulation to a real animal under the same experimental conditions. While the results will depend on the ability of our subjects to accurately judge how realistic the emotional states and behavior of the pet are, we feel that most people have sufficient common knowledge to perform this task.

4.2.1.1. Selection of Subjects

Participants in the evaluation were recruited by email from a list of undergraduates that is maintained by the Computer Science Department at Texas A&M University. We recruited 21 subjects from this pool. The ages of the subjects were in the range of 18-26 years old. Most of the participants were first year undergraduates. This source was chosen to reduce the bias due to the background knowledge of the users within the sample. Users with specialized backgrounds might evaluate the simulation according to their own field. For example, users with a sociology background will look at the social side of the pet, people with an HCI background will look at the usability side, and people with biology background might compare the pet to a real animal. In order to diminish this effect we selected first year undergraduates who do not have specific background knowledge tied to any field.

4.2.1.2. Experimental Procedure

Participants were asked to meet with the principle investigator for a period of two and one-half hours in a computer science lab at the Texas A&M campus. The participants were first given a fifteen-minute introduction to the system. In this introduction, they were notified that their responses would be used to further enhance the system, not as a measure of how good the system is. This way, the users were encouraged to give constructive criticism of the system, while avoiding assessing the system in an unduly positive fashion because they were either trying to be nice or were impressed by the interface.

Participants were handed instruction sheets, which walked them through different scenarios of the simulation, showing them the different aspects of the pet's behavior. Eventually, they were asked to fill out a questionnaire about what they observed in the simulation. While answering the questionnaires, the subjects were asked not to write their names or mark the papers in any way that might be used to identify them at a later time period, to ensure the anonymity of response.

The protocol described above was repeated for four different versions of the system. (1) We evaluated a simulation where the pet produced a set of random emotions and random behaviors. This version provided a baseline for other experiments. Sometimes users might be impressed by the pictures or sound effects; or the fact that the pet in the pictures reacts at all to the user's actions might influence some users to answer some questions with positive feedback. (2) We evaluated a non-random version of the system, but with no fuzzy logic or learning, using a crisp interval mapping instead of fuzzy mapping and constant probabilities for expectation values. (3) A version with fuzzy logic but no learning was evaluated to observe the effect of fuzzy logic on the system. (4) A simulation with both fuzzy logic and learning was then evaluated to determine the advantage of making the model adaptive.

Due to the limited sample size, we carried out the experiments so that each user would use and answer questions on all four versions of the system. To eliminate the effect of users making direct comparisons with other versions they have observed in the course of the experiment, we employed a counter-balanced Latin square design, where the order of presentation of the different models was shuffled for the different users.

4.3 Results

The questionnaires were collected and analyzed for 21 subjects. The questions in the questionnaire were designed to evaluate the different elements of the model. In this section, we will present the quantitative results, and we also give some of the users' informal comments.

4.3.1 Intelligence

To explore how users perceive PETEEI's intelligence in the four models, we asked the users to rate the intelligence of PETEEI based on certain definitions, specifically goal-oriented behavior and adaptability. We formulated the questions as follows:

```
The concept of intelligence has been defined and redefined
many times. For the purpose of our experiment we will
evaluate intelligence from two different perspectives.

A. Intelligence was also defined as exerting behaviors or
   actions that are goal oriented.

       Do you think that PETEEI has this form of
       intelligence?
         ☐ Yes               ☐ No

   If yes, rate your answer in the scale of zero to ten.
   (0| the pet is has only minor goal-directed
   intelligence→ 10| the pet is very goal-directly
   intelligent). Explain your rating.
```

B. Another definition to intelligence is the ability of the agent or the subject to adapt to a certain environment or situation.

Do you think that PETEEI has this form of intelligence?

☐ Yes ☐ No

If yes, rate your answer in the scale of zero to ten. (0| the pet is has only minor adaptability→ 10| the pet is very adaptable). Explain your rating.

C. Overall how would you rate the pet's intelligence using the three criteria above? Explain your answer.(0| not intelligent → 10|very intelligent).

We used the following formula to calculate the statistical significance of the binary answers above using a 95% confidence interval:

$$\pm 1.96\sqrt{\frac{p*(1-p)}{n}}$$

where n is the sample size, p is the percentage of Yes answers and *1.96* represents the 2-sided z-score for *95%* confidence. For answers on a scale from 0-10, we calculated the standard error of the mean for a sample size of 21. This measure was calculated using the following formula:

$$\bar{x} \pm \frac{\sigma}{\sqrt{n}}$$

where $\bar{x}$ is the mean, the σ is the standard deviation and n is the size of the sample, which is 21 in this case.

Table 4 shows the means, standard errors and confidence intervals of the responses. Figure 10 shows these results graphically. The figure depicts the four models in different gray shades. The four models depicted in the figure are the random model, non-fuzzy/non-learning, fuzzy/non-learning, and fuzzy/learning model, respectively. The figure shows the mean values of the models. The standard errors are depicted by the error bars.

The intelligence ratings of the random model were 0.14 for goal-oriented behavior, 0.33 for adaptability, and 1.14 overall. One user rated the pet's intelligence in the random model as 2 out of 10, and they explained their answer by saying "PETEEI knew about his surroundings, like he knew if there was water around him or not." These scores for the random model establish a baseline for the other responses. In contrast, by using the non-fuzzy/non-learning model, the level of intelligence increased a little bit (significant for overall intelligence, Question C). Interestingly, the addition of fuzzy logic did not further increase the intelligence ratings based on these definitions. However, the addition of learning to the model brought about a much more significant increase in the intelligence ratings. The overall intelligence went from around 3.05 to 7.05. A significant increase was also noted for questions A and B on the goal-oriented and adaptive aspects of intelligence specifically.

4.3.2 Learning

To gain deeper insight into the impact of the adaptive component of FLAME on the believability of the simulation, we asked subjects more directed questions about the learning behavior of PETEEI. These questions examined subjects' perceptions of PETEEI's ability to adapt to its environment and the user, which requires learning about event probabilities, user tendencies, good and bad actions based on user feedback, and associations between events/objects and emotional reactions[2].

The questions that were asked of the subjects about the learning behavior of the model in the simulation were:

```
Learning can be measured in many different aspects.


A. A subject x can learn about the other people's behavior
   to know what to expect and from whom.

   Do you think PETEEI learns about you?
      Yes  ☐      No  ☐


   If yes, rate your answer in the scale of zero to ten.
   (0| the pet leans only a few things about me → 10| the
   pet learns a lot about me). Explain your rating.


B. A subject could learn more about the environment to
   plan for his/her actions.

      Do you think PETEEI learns about its environment?
          ☐ Yes        ☐ No


   If yes, rate your answer in the scale of zero to ten.
   (0| the    pet leans only a few things about the
   environment→ 10| the pet learns a lot about the
   environment). Explain your rating.


C. A subject could learn how to evaluate some actions as
   good or bad according to what people say or according
   to his own beliefs.

   Do you think PETEEI learns about good and bad actions?
         ☐ Yes                      ☐ No


   If yes, rate your answer in the scale of zero to ten.
   (0| the pet leans about only a few actions→ 10| the
   pet learns to assess all its actions). Explain your
   rating.


D. Overall how would you rate the pet's learning using the
   four criteria listed above? Explain your answer.(0|
   does not learn much→ 10|learns a lot).
```

[2] We note that for these experiments, an older version of the FLAME model was used (Seif El-Nasr et al. 1999b). While most of the learning algorithms were the same as described in section 3 above, the learning of action values (i.e., how much the actions please or displease the user) was done using small constant multipliers to update estimates of action values. This could make a slight difference in the relative assessment of the quality of actions, but the overall effect should be the same – that actions leading to positive feedback would become known as pleasing actions and vice versa.

The results, shown in Table 5 and Figure 11, revealed how learning was perceived by users. Clearly, learning was only perceived when the learning component of the model was used. The responses for all other versions of the system were under 2.0. There were some minor differences that can be noted from the table among these different models; however, the confidence intervals overlap to a great extent, leading us to believe that users did not perceive significant differences in these models. In contrast, the learning model was observed to increase the learning perceived in all questions. In question A (learning about the user), there was an increase to 7.7; in question B (learning about the environment), there was an increase to 5.95; in question C (learning from feedback), there was an increase to 7.67, and in question D (overall) there was an increase in the ratings to 7.95. Thus, the learning component was responsible for the apparent adaptability of the system, and this is a reflection of the different types of learning which are necessary for a believable model of emotions.

4.3.3 Behavior

Finally, the users were asked to rate PETEEI in terms of how convincing the pet's behavior was. The users were told not to rate it in terms of animation quality or facial expression, but to concentrate on the behavioral and emotional aspects. We analyzed users' answers and we present the results in Table 6.

As the table above shows, the random model did not convey a realistic or convincing behavior to the user since the rating was on average 1.0. In contrast, the introduction of the non-fuzzy/non-learning model improved this measure by about 3.3 units on average. The fuzzy/non-learning model also improved this measure by 1.1 units. The fuzzy/learning model further improved this measure by 2.6 to an average of 8.1 units. All of these increases are likely significant.

4.3.4 Summary

We conclude that the introduction of learning improved the system and created a closer simulation of the behavior of a real pet. While fuzzy logic did not contribute significantly in the user evaluations, it provided a better means of modeling emotions due to its qualitative and quantitative expressiveness. Learning, on the other hand, was found to be most important for creating the appearance of intelligence and believability. Even though the pet was not animated, the users all agreed that the model behind the interface could serve further as a basis to simulate more believable characters. As a matter of fact, one of our users noted that, "I like this model better than Petz [a commercial product that simulates believable pets], because this model shows me more about how the pet learns and displays his affections to me, while in Petz the dogs just play around without any emotions, or if there were some, I was not able to determine what emotions the pets were feeling and how they learned from me."

5 Discussion

FLAME could be used as a computational model of emotions to enhance a variety of different computer interfaces or interactive applications. For instance, FLAME could be used to implement a believable agent in animated character applications (Thomas and Johnston 1981), such as interactive theatre productions or role-playing video games. These often involve simulation of synthetic characters which interact with the user. Typically such applications are designed using a scenario-driven technique, in which the animator must anticipate and script responses to all possible sequences of events. However, there has been recent interest in using intelligent agents to generate realistic behavior autonomously (Maes 1995, Blumberg 1996), and a model of emotions, such as FLAME, could be used to enhance the believability of such character animations (Seif El-Nasr et al. 1999a). The utility of a computational model of emotions can also be foreseen in educational software (Elliot et al. 1997). The incorporation of more synthetic and emotionally expressive characters has been shown to enhance the learning process in children (Lester et al. 1997). FLAME can also be used as the basis for producing a more responsive tutor agent in training simulations (Rickel and Johnson 1999). Finally, a model of emotions like FLAME could potentially be used to enhance human-computer interfaces (Picard 1997). Much of the work in this area has focused on inferring the internal states of the user based on what events he/she has seen on the screen and what actions he/she has taken (Benyon and Murray 1993). FLAME could be employed as the basis of a user model to also track the user's likely emotional state according to these inputs (in addition to implementing a believable agent for display (Koda and Maes 1996)).

Before porting FLAME to other applications, however, we would need to address some of its limitations. Aside from specifying the new goals of the agent, some parameters used in the model might need to be adjusted. Through our discussion of the model, we have introduced many different parameters, including the decay constant of negative and positive emotions, the impact β of the mood on expectations, the tolerance degree that measures the closeness of emotions, the number of pervious time periods for calculating the mood, etc. These parameters might need to be set to specific values before linking the model to a different application. A simple approach would be to implement an experimental version of the system and then use trial and error over some range of parameter values to determine the optimal settings. A more sophisticated approach would need to be used in cases where the values are not independent.

There are many ways in which FLAME could be extended. For example, the model, as described above, does not incorporate personality. Simulating virtual characters involves more than just simulating emotional behavior; personality is regarded as one of the most important factors that differentiates people (Nye and Brower 1996), and is thus one of the most important features that can enhance believability of animated characters (Thomas Johnston 1981). Several researchers within the social agents' community incorporated personality in their model (Loyall and Bates 1997, Blumberg 1996, Rousseau 1996). For example, D. Rousseau and B. Hayes-Roth simulated agents with different personalities by using rules to define different personality traits (Rousseau 1996, and Rousseau and Hayes-Roth 1997). The personality traits they developed include: introverted, extroverted, open, sensitive, realistic, selfish, and hostile (Rousseau 1996). They described each trait in terms of the agent's inclination or focus, which determined aspects of the agent's behavior (Rousseau and Hayes-Roth 1997).

Personality theory has been addressed within many disciplines. Many approaches have been proposed to account for individual differences, including evolutionary constructs, biological and genetic substrates, affective and cognitive constructs, and the self and the other theory (Revelle 1995). Recently, strong evidence of the role of moods, motivations, and experience in shaping personality and individual differences was found (Revelle 1993). Therefore, simulating personality may involve exploring the relationship with other related processes, such as learning, emotions, and cognition. Incorporating personality into FLAME would be a difficult but important task. One interesting idea is that we might be able to account for some personalities by manipulating the parameters of the model. For example, a very optimistic person may have a much larger mood factor β for positive events than negative events. Nevertheless, additional research needs to be done to explore the relationships among moods, experience, emotions and personality.

Finally, we note that in many applications an agent rarely acts by itself; rather, it is often a member of a group of agents that interact with each other to accomplish various tasks (e.g., software agents) (Huhns and Singh 1998). To simulate a group of agents that interact with each other, we would have to extend FLAME to incorporate some of the social agents concepts. FLAME was designed to model the interaction between one agent and a user. For an agent to interact with other agents, the architecture will have to be extended to account for multiple mental models, as was discussed by Elliot (1992). We will also have to add other aspects that psychologists have identified to influence emotions and social interactions, including self-evaluation, self-perception, self-awareness, self-realization and self-esteem (Nye and Brower 1996, and Goleman 1995). In essence, the problem will then be extended from modeling emotions to modeling social behavior and interactions.

6 Conclusion

In this paper, we have described a new computational model of emotions called FLAME. FLAME is based on an event-appraisal psychological model and uses fuzzy logic rules to map assessments of the impact of events on goals into emotional intensities. FLAME also includes several inductive algorithms for learning about event expectations, rewards, patterns of user actions, object-emotion associations, etc. These algorithms can enable an intelligent agent implemented with FLAME to adapt dynamically to users and its environment. FLAME was used to simulate emotional responses in a pet, and various aspects of the model were evaluated through user-feedback. The adaptive components were found to produce a significant improvement in the believability of the pet's behavior. This model of emotions can potentially be used to enhance a wide range of applications, from character-based interface agents, to animated interactive video games, to the use of pedagogical agents in educational software. These could all benefit from the generation of autonomous believable behavior that simulates realistic human responses to events in real-time. There are a number of ways in which FLAME could be extended, such as taking into account personality, self-esteem, social behavior (multi-agent interactions), etc. Nonetheless, the adaptive capabilities of FLAME represent a significant improvement in our ability to model emotions, and can be used as the basis to construct even more sophisticated and believable intelligent systems.

REFERENCES

J. Armony, J. Cohen, D. Servan-Schreiber, and J. LeDoux. (1995). An Anatomically Constrained Neural Network Model of Fear Conditioning. *Behavioral Neuroscience*, 109 (2), 240-257.

R. M. Baecker, J. Grudin, W. A. S. Buxlon, and S. Greenberg, Eds. (1995). *Readings in Human-Computer Interaction: Toward the Year 2000, 2nd. Ed*. San Francisco, CA: Morgan Kaufmann Publishers.

J. Bates. (1992). The Role of Emotion in Believable Agents. *Communications of the ACM*, 37(7), 122-125.

J. Bates, A. B. Loyall, and W. S. Reilly. (1992a). An Architecture for Action, Emotion, and Social Behavior. School of Computer Science, Pittsburgh, PA: Carnegie Mellon University, Technical Rep. CMU-CS-92-144.

J. Bates, A. B. Loyall, and W. S. Reilly. (1992b). Integrating Reactivity, Goals and Emotion in a Broad Agent. School of Computer Science, Pittsburgh, PA: Carnegie Mellon University, Technical Rep. CMU-CS-92-142.

D. R. Benyon and D. M. Murray. (1993). Adaptive Systems; from intelligent tutoring to autonomous agents. *Knowledge-based Systems*, 6 (3).

C. Breazeal and Brian Scassellati. (to appear). Infant-like Social Interactions between a Robot and a Human Caretaker. To appear in *Special Issue of Adaptive Behavior on Simulation Models of Social Agents.*

R. Brooks, C. Breazeal, M. Marjanovic, B. Scassellati, and M. Williamson. (to appear). The Cog Project:Building a Humanoid Robot. To appear in *A Springer-Verlag Lecture Notes in Computer Science*.

B. M. Blumberg, P. M. Todd, P. Maes. (1996). No Bad Dogs: Ethological Lessons for Learning. *In: From Animals To Animats, Proceedings of the Fourth International Conference on the Simulation of Adaptive Behavior*.

B. M. Blumberg (1996). Old Tricks, New Dogs: Ethology and Interactive Creatures. Ph.D. Thesis, MIT media Lab, Cambridge, MA.

R. C. Bolles and M. S. Fanselow. (1980). A Perceptual Defensive Recuperative Model of Fear and Pain. *Behavioral and Brain Sciences*, 3, 291-301.

G. H. Bower and P. R. Cohen. (1982). Emotional Influences in Memory and Thinking: Data and Theory. *Affect and Cognition*. M. Clark and S. Fiske. London, Eds.: Lawrence Erlbaum Associated Publishers, 291-331.

A. R. Damasio. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain.* New York: G.P. Putnam.

R. Descartes. (1989). *The Passions of the Soul*. Trans. Stephen Voss. Cambridge: Hackett Publishing Company.

Kerstin Dautenhahn. (1998). The Art of Designing Socially Intelligent Agents: Science, Fiction and the Human in the Loop. Applied Artificial Intelligence journal, Special Issue on "Socially Intelligent Agents" 12 (7-8), 573-619.

M. Domjan (1998). *The principles of learning and behavior* (4th ed.). Boston: Brooks/Cole Publishing Co.

P. Ekman. (1992). An Argument for Basic Emotions. *Cognition and Emotion*, London, Lawrence Erlbaum Associated Publishers, 169-200.

C. Elliot. (1994). Research Problems in the use of a Shallow Artificial Intelligence Model of Personality and Emotion. *AAAI '94*.

C. Elliot. (1992). The Affective Reasoner: A Process model of emotions in a multi-agent system. Institute for the Learning Sciences, Evanston, IL: Northwestern University, Ph.D. Thesis.

C. Elliott, J. Rickel, and J. Lester (1997). Integrating Affective Computing into Animated Tutoring Agents. Fifteenth International Joint Conference on Artificial Intelligence '97 — Animated Interface Agents Workshop, 113-121.

C. Elliot and J. Brzezinski. (1998). Autonomous Agents as Synthetic Characters. *AI Magazine*, American Association for Artificial Intelligence, 13-30.

A. P. Fiske (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychology Review,* 99, 689-723.

J. P. Forgas (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin*, 117, 39-66.

J. P. Forgas (1994). Sad and guilty?: Affective influences on the explanation of conflict in close relationships. *Journal of Personality and Social Psychology*, 66, 56-68.

H. Gardner. (1983). *Frames of Mind*. New York: Basic Books.

D. Goleman. (1995). *Emotional Intelligence*. New York: Bantam Books.

M. Huhns and M. P. Singh. Eds. (1998). *Reading In Agents*. Morgan Kaufmann Publishers, Inc., CA.

K. Inoue, K. Kawabata, and H. Kobayashi. (1996). On a Decision Making System with Emotion. *IEEE Int. Workshop on Robot and Human Communication*, Tokyo, Japan, 461-465.

C. E. Izard. (1977). *Human Emotions*. New York & London: Plenum Press.

W. James. (1884). What is an Emotion? *Mind,* 9, 188-205.

L. P. Kaelbling, M. L. Littman, and A. W. Moore. (1996). Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, 4, 237-285.

T. Koda and P. Maes (1996). Agents with Faces: The Effect of Personification. *Proceedings of 5th International Workshop on Robot and Human Communication*, 189-194.

N. R. Jennings, K. Sycara, and M. Wooldridge (1998). A Roadmap of Agent Research and Development. *Autonomous Agent and Multi-Agent Systems*, 1(1), 7-38.

J. Kim and J. Yun Moon. (1998). Designing towards emotional usability in customer interfaces – trustworthiness of Cyber-banking system interfaces. *Interacting with Computers: The Interdisplincary Journal of Human-Computer Interaction*, 10(1).

B. Kosko. (1997). *Fuzzy Engineering*. Prentice Hall, USA.

A. Konev, E. Sinisyn, and V. Kim. (1987). Information Theory of Emotions in Heuristic Learning. *Zhurnal Vysshei Nervnoi Deyatel'nosti imani*, 37 (5), 875-879.

J. LeDoux. (1996). *The Emotional Brain*, New York: Simon & Schuster.

J. Lester, S. Converse, S. Kahler, T. Barlow, B. Stone, and R. Bhogal. (1997). The Persona Effect: Affective Impact of Animated Pedagogical Agents. in *Proc. CHI '97 Conference*, Atlanta, GA.

A. B. Loyall and J. Bates. (1997). Personality-Based Believable Agents That Use Language. *Proc. of the First Autonomous Agents Conference*, Marina del Rey, CA.

P. Maes. (1994). Agents that Reduce Work and Information Overload. *Communications of ACM*, 37(7), 37-45.

P. Maes. (1995). Artificial Life Meets Entertainment: Lifelike Autonomous Agents. *Communications of the ACM Special Issue on Novel Applications of AI*, 38 (11), 108-114.

E. H. Mamdani and S. Assilian. (1975. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Machine Studies*, 7 (1).

M. A. Mark (1991). *The VCR Tutor: Design and Evaluation of an Intelligent Tutoring System*. Master's Thesis, University of Saskatchewan, Saskatoon, Saskatchewan.

E. Masuyama. (1994). A Number of Fundamental Emotions and Their Definitions. *IEEE International Workshop on Robot and Human Communication*, Tokyo, Japan, 156-161.

G. A. Miller. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits On Our Capacity for Processing Information. *Psychological Review*, 63, 81-97.

M. Minsky. (1986). *The Society of the Mind*. New York: Simon and Schuster.

T. M. Mitchell. 9196). *Machine Learning*, New York: McGraw-Hill Co..

H. Mizogouhi, T. Sato, and K. Takagi. (1997). Realization of Expressive Mobil Robot. *Proc. IEEE Int. Conf. of Robotics and Automation*, Albuquerque, NM, 581-586.

H. S. Nwana. (1996). Software Agents: An Overview. *Knowledge Engineering Review*, 11 (3), 205-224.

J. L. Nye and A. M. Brower (1996). *What's Social About Social Cognition?* London: Sage Publications.

A. Ohman. (1994). *The Psychophysiology of Emotion; Evolutionary and Non-conscious Origins*. London, England: Lawrence Erlbaum Accociates, Inc..

A. Ortony, G. Clore, and A. Collins. (1988). *The Cognitive Structure of Emotions*. Cambridge: Cambridge University Press.

R. Pfeifer. (1988). Artificial Intelligence Models of Emotions. *Cognitive Perspectives on Emotion and Motivation*, 287-320.

R. W. Picard and J. Healy. (1997). Affective Wearables. *Proc. of IEEE Conf. on Wearables*, 90-97.

R. W. Picard. (1995). Affective Computing. Cambridge, MA: MIT Media Lab, Technical Rep. No. 221.

R. W. Picard. (1997). *Affective Computing*, Cambridge, MA: MIT press.

D. D. Price, J. E. Barrell, and J. J. Barrell. (1985). A Quantitative-Experiential Analysis of Human Emotions. *Motivation and Emotion*, 9 (1).

W. S. Reilly. (1997). A Methodology for Building Believable Social Agents. *The First Autonomous Agent's Conference*, Marina del Rey, CA.

W. S. Reilly. (1996). Believable Social and Emotional Agents, School of Computer Science, Pittsburgh, PA: Carnegie Mellon University, Ph.D. Thesis CMU-CS-96-138.

W. S. Reilly and J. Bates. (1992). Building Emotional Agents. Pittsburgh, PA: Carnegie Mellon University, Technical Rep. CMU-CS-92-143.

R. A. Rescorla. (1988). Pavlovian Conditioning: It's Not What You Think It Is. *American Psychologist*, 43(3), 151-160.

R. A. Rescorla. (1991). Associative Relations in Instrumental Learning: The Eighteenth Bartlett Memorial Lecture. *The Quarterly Journal of Experimental Psychology*, 43B (1), 1-25.

W. Revelle (1993). Individual differences in personality and motivation: Non-cognitive determinants of cognitive performance. In A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness and control: A tribute to Donald Broadbent*, 346-373. Oxford: Oxford University Press.

W. Revelle (1995). Personality Processes. Annual Review of Psychology, 46, 295-328.

J. Rickel and W. L. Johnson (1999). Animated Agents for Procedural training in Virtual reality: Preception, Cognition, and Motor Control. To appear in *Journal of Applied Artificial Intelligence*.

D. R. Riso and R. Hudson (1996). *Personality Types*, USA: Houghton Mifflin.

I. J. Roseman, P. E. Jose, and M. S. Spindel. (1990). Appraisals of Emotion-Eliciting Events: Testing a Theory of Discrete Emotions. *Journal of Personality and Social Psychology*, 59 (5), 899-915.

D. Rousseau and B. Hayes-Roth. (1997). Interacting with Personality-Rich Characters Stanford Knowledge Systems Laboratory Report KSL-97-06.

D. Rousseau. (1996). Personality in computer characters. In Working Notes of *the AAAI-96 Workshop on AI/ALife*, AAAI Press, Menlo Park, CA, 1996.

S. Russell and P. Norvig. (1995). *Artificial Intelligence A Modern Approach*, Upper Saddle River, NJ: Prentice-Hall Inc.

K. R. Scherer. (1993). Studying the Emotion Antecedent Appraisal Process: An Expert System Approach. *Cognition and Emotion*, 7, 325-55.

R. Schumacher and M. Velden. (1984). Anxiety, Pain Experience and Pain Report: A Signal-detection Study. *Perceptual and Motor Skills*, 58, 339-349.

M. Seif El-Nasr. (1998). Modeling Emotional Dynamics in Intelligent Agents. Master's thesis, Computer Science Department, Texas A&M University.

M. Seif El-Nasr, T. Ioerger, J. Yen, F. Parke, and D. House. (1999a). Emotionally Expressive Agents. *Proceedings of Computer Animation Conference 1999,* Geneva, Switzerland.

M. Seif El-Nasr, T. Ioerger, and J. Yen (1999b). PETEEI: A PET with Evolving Emotional Intelligence. *Proceedings of the Third International Conference on Autonomous Agents*, Seattle, Washington.

T. Shibata, K. Inoue, and Robert Irie. (1996). Emotional Robot for Intelligent System -Artificial Emotional Creature Project-. *IEEE Int. Workshop on Robot and Human Communication*, Tokyo, Japan, 466-471.

T. Shibata, K. Ohkawa, and K. Tanie. (1996). Spontaneous Behavior of Robots for Cooperation - Emotionally Intelligent Robot System-. *Proc. of IEEE Int. Conf. on Robotics and Automation*, Japan, 2426-2431.

T. Shiida. (1989). An Attempt to Model Emotions on a Machine. *Emotion and Behavior: A System Approach,* 2, 275-287.

H. Simon. (1967). Motivational and Emotional Controls of Cognition. *Psychological Review*, 74 (1), 29-39.

H. Simon. (1996). *The Science of the Artificial*, Cambridge, MA: MIT Press.

L. Suchman. (1981). *Plans and Situated Actions*, Cambridge University Press.

S. Sugano and T. Ogata. (1996). Emergence of Mind in Robots for Human Interface - Research Methodology and Robot Model -. *Proc. International Conference on Robotics and Automation IEEE*, Japan, 1191-1198.

H. Takagi and M. Sugeno. (1985). Fuzzy identification of systems and its application to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics*, 15 (1).

S. Tayrer, Ed. (1992). *Psychology, Psychiatry and Chronic Pain*. Oxford, England: Butterworth Heinemann.

F. Thomas and O. Johnston. (1981). *The Illusion of Life*, New York: Abbeville Press.

J. Velasquez. (1997). Modeling Emotions and Other Motivations in Synthetic Agents. *Proceedings of the AAAI Conference 1997*, Providence, RI, 10-15.

J. Yen. (1999). Fuzzy Logic: A modern Perspective. *IEEE Transactions on Knowledge and Data Engineering,* 11 (1), 153-165.

J. Yen and R. Langari. (1999). *Fuzzy Logic: Intelligence, Control and Information*, Upper Saddle River, NJ: Prentice Hall.

L. A. Zadeh. (1965). Fussy Sets. *Information and Control,* Vol. 8.

FIGURES AND TABLES

Figures:

Tables**:**

Table 1. Rules for Generation of Emotions

Table 2. Calculating Intensities of Emotions

Table 3. Calculating the Intensity of Motivational States in PETEEI**.**

Table 4. Intelligence Ratings of PETEEI's intelligence with respect to goal-directed behavior (question A), adaptation to the environment and situations happening around it(Question B), and overall impression of PETEEI's intelligence.

Table 5. User's evaluations of aspects of learning in various versions of PETEEI: A) about users, B) about environment, C) about good or bad actions, and D) overall impression that PETEEI learns.

Table 6. User's ratings of how convincing the behavior of the pet in PETEEI was**.**

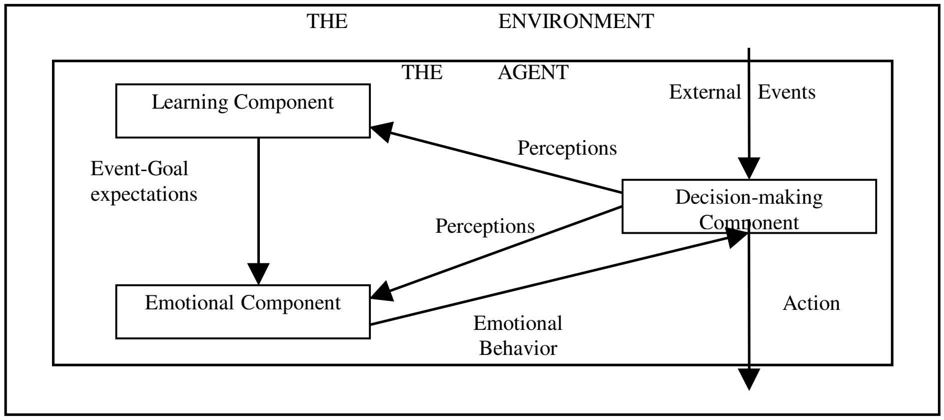

THE ENVIRONMENT
THE AGENT
Learning Component
External Events
Perceptions
Event-Goal expectations
Decision-making Component
Perceptions
Emotional Component
Action
Emotional Behavior

Figure 1

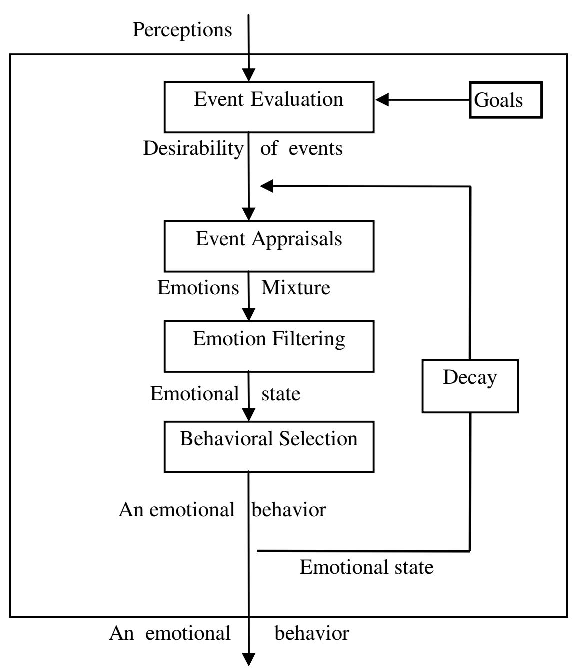
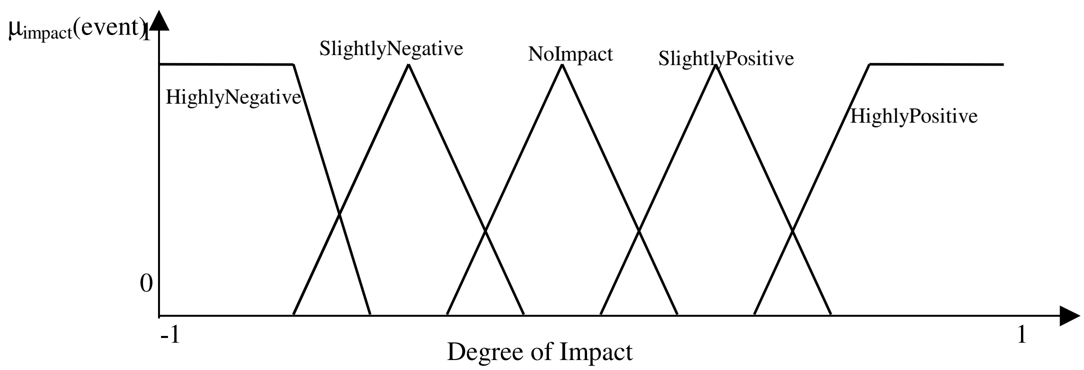
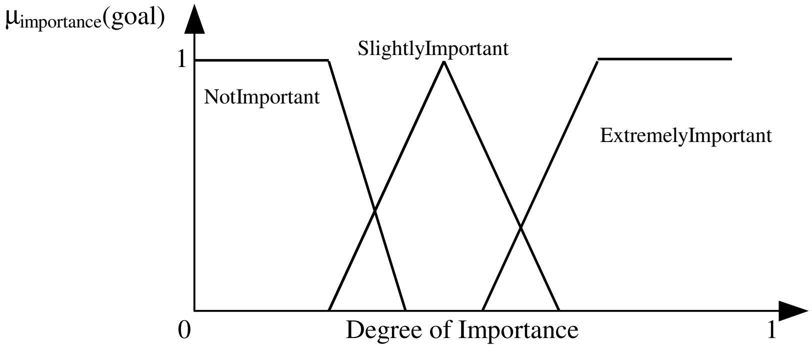

Perceptions
Event Evaluation
Goals
Desirability of events
Event Appraisals
Emotions Mixture
Emotion Filtering
Emotional state
Decay
Behavioral Selection
An emotional behavior
Emotional state
An emotional behavior

Figure 2



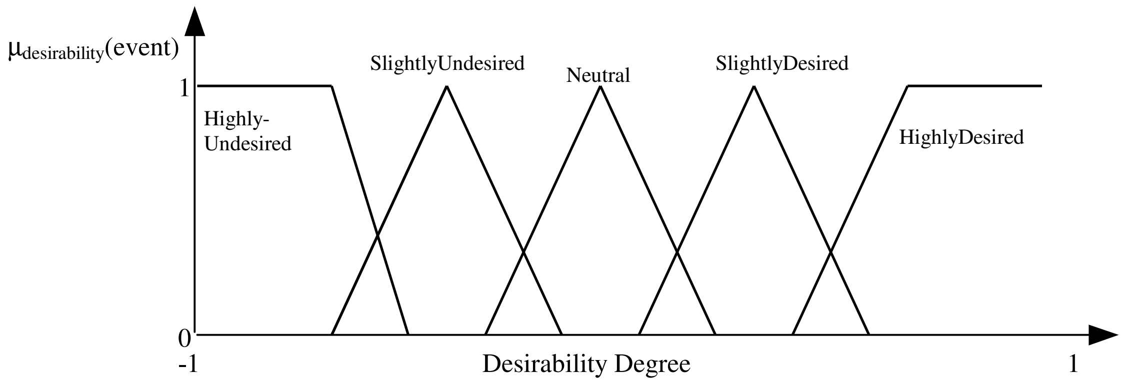
μimpact(event)
1
0
SlightlyNegative
NoImpact
SlightlyPositive
HighlyNegative
HighlyPositive
-1
1
Degree of Impact

Figure 3



μimportance(goal)
1
SlightlyImportant
NotImportant
ExtremelyImportant
0
Degree of Importance
1

Figure 4



μdesirability(event)
1
0
-1
1
Highly-
Undesired
SlightlyUndesired
Neutral
SlightlyDesired
HighlyDesired
Desirability Degree

Figure 5

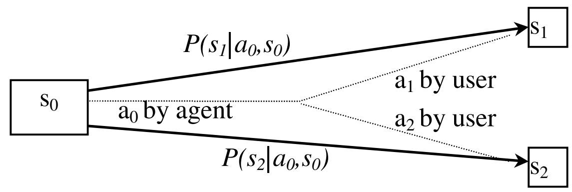
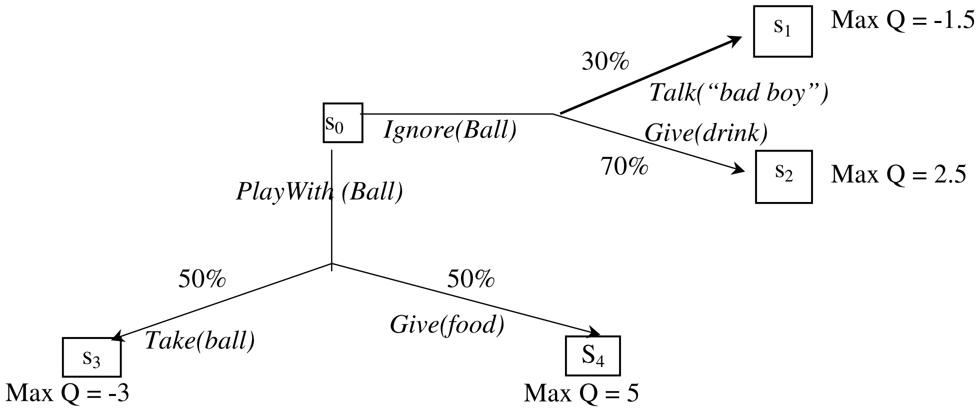


Figure 6



s1
Max Q = -1.5
30%
Talk("bad boy")
s0
Ignore(Ball)
Give(drink)
70%
s2
Max Q = 2.5
PlayWith (Ball)
50%
50%
Give(food)
Take(ball)
s3
S4
Max Q = -3
Max Q = 5

Figure 7

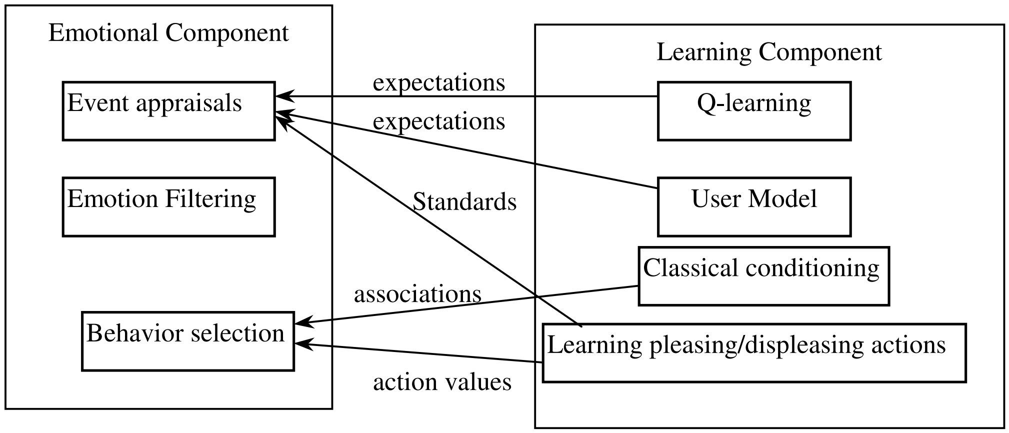

Emotional Component
Event appraisals
Emotion Filtering
Behavior selection
Learning Component
Q-learning
User Model
Classical conditioning
Learning pleasing/displeasing actions
expectations
expectations
Standards
associations
action values

Figure 8

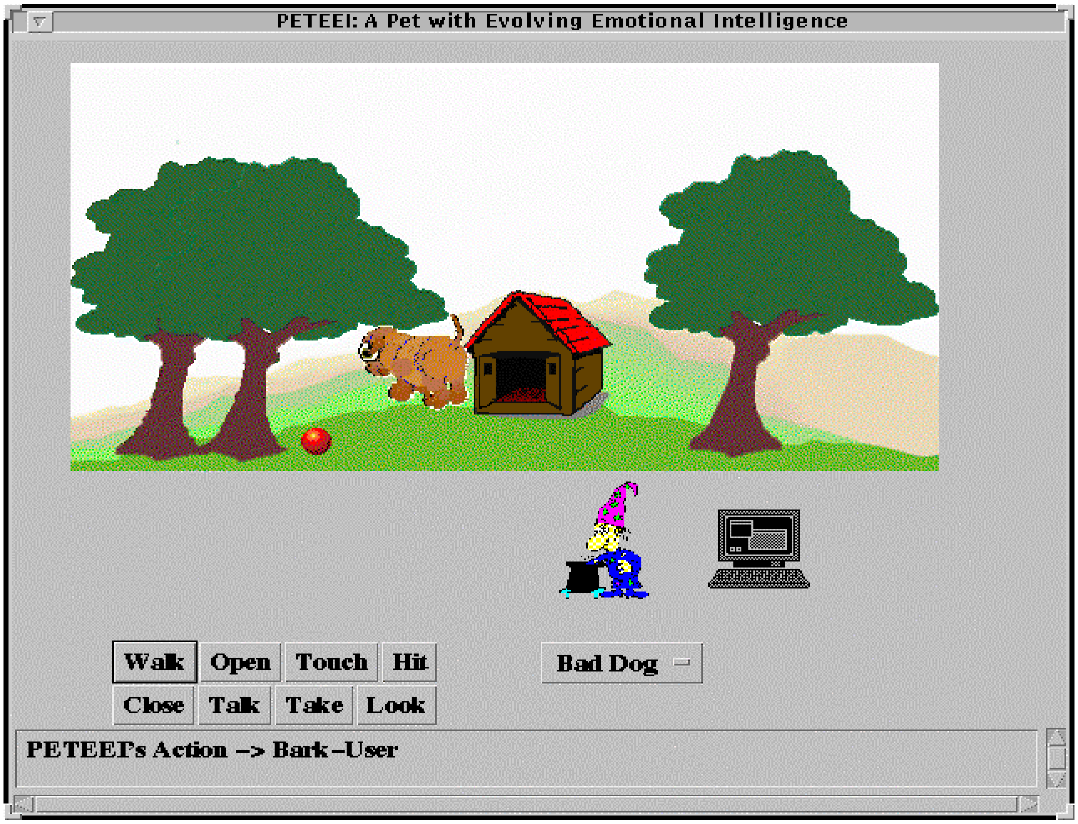

PETEEI: A Pet with Evolving Emotional Intelligence
Walk
Open
Touch
Hit
Close
Talk
Take
Look
Bad Dog
PETEEI's Action -> Bark-User

Figure 9 (a)

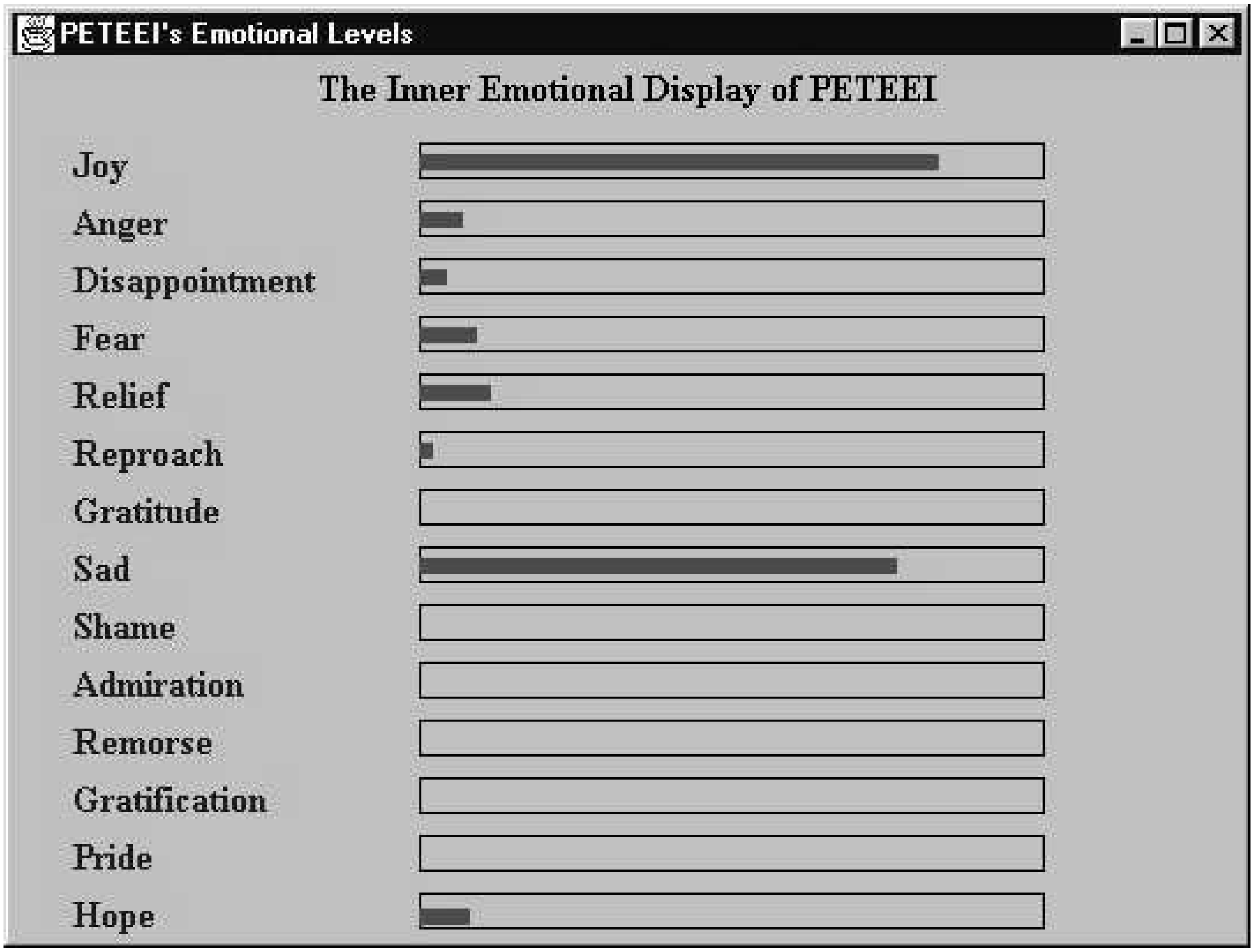

PETEEI's Emotional Levels
The Inner Emotional Display of PETEEI
Joy
Anger
Disappointment
Fear
Relief
Reproach
Gratitude
Sad
Shame
Admiration
Remorse
Gratification
Pride
Hope

Figure 9(b)

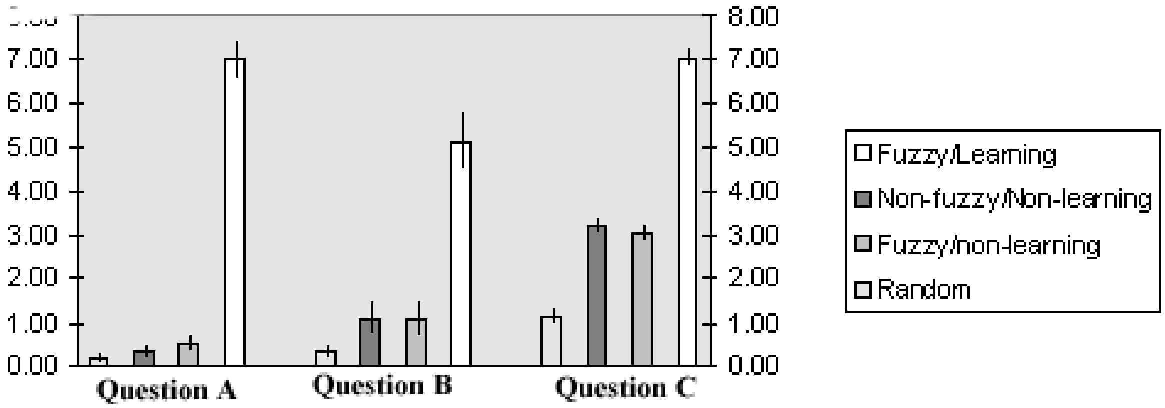

8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
Question A
Question B
Question C
Fuzzy/Learning
Non-fuzzy/Non-learning
Fuzzy/non-learning
Random

Figure 10

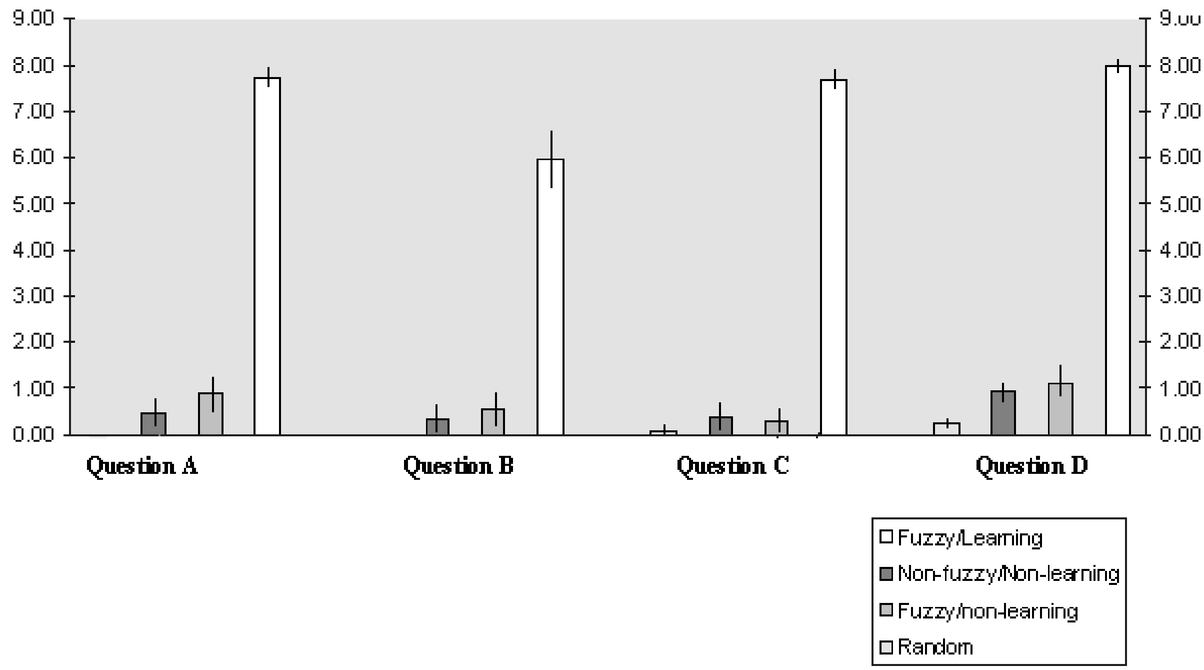

9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
Question A
Question B
Question C
Question D
Fuzzy/Learning
Non-fuzzy/Non-learning
Fuzzy/non-learning
Random

Figure 11

Emotion	Rule
Joy	occurrence of a desirable event
Sad	occurrence of an undesirable event
Disappointment	occurrence of a disconfirmed desirable event
Relief	occurrence of a disconfirmed undesirable event
Hope	occurrence of an unconfirmed desirable event
Fear	occurrence of an unconfirmed undesirable event
Pride	action done by the agent and is approved by standards
Shame	action done by the agent and is disapproved by standards
Reproach	action is done by the other and is not approved by the agents standards
Admiration	action done by the other and is approved by the agents standards
Anger	Complex emotion; sad + reproach
Gratitude	Complex emotion; joy + admiration
Gratification	Complex emotion; joy + pride
Remorse	Complex emotion; sad + shame

Table 1

Emotion	Formula for Intensity
Joy	$Joy = (1.7 \times expectation^{0.5}) + (-0.7 \times desirability)$
Sadness	$Sad = (2 \times expectation^{2}) - desirability$
Disappointment	$Disappointment = Hope \times desirability$
Relief	$Relief = Fear \times desirability$
Hope	$Hope = (1.7 \times expectation^{0.5}) + (-0.7 \times desirability)$
Fear	$Fear = (2 \times expectation^{2}) - desirability$

Table 2

Motivational State	**Intensity is function of…**
Pain	Number of Hits by user, the intensity of the hits
Tired	Amount of movement, time last slept
Thirst	frequency of movement, intensity of movement (running is higher than walking ..etc.), time last drank, quantity that was drunk
Hunger	intensity of movement, frequency of movement, time last ate, the quantity of food eaten

Table 3

Model	Question A		Question B		Overall	
	mean	error	mean	Error	mean	error
Random	0.14	0.04-0.25	0.33	0.16-0.51	1.143	1-1.29
Non-fuzzy/non-learning	0.3	0.19-0.485	1.1	0.74-1.45	3.19	3.03-3.35
Fuzzy/non-learning	0.5	0.33-0.71	0.5	0.73-1.46	3.05	2.86-3.24
Fuzzy/learning	7	6.57-7.43	5.1	4.49-5.79	7.05	6.86-7.24

Table 4

Model	Question A		Question B		Question C		Question D	
	mean	error	mean	error	mean	error	mean	error
Random	0	0 - 0	0	0 - 0	0.1	0-0.19	0.24	0.14 - 0.33
Non-fuzzy/ non-learning	0.5	0.18-0.77	0.3	0.05-0.62	0.4	0.08-0.68	0.91	0.67 - 1.14
Fuzzy/non-learning	0.9	0.45-1.26	0.5	0.16-0.89	0.29	0.05-0.53	1.14	0.8-1.48
Fuzzy/learning	7.7	7.49-7.93	5.95	5.35-5.56	7.67	7.46-7.88	7.95	7.78-8.13

Table 5

Model	Behavior	
	mean	interval
Random	1.0	0.83-1.17
Non-fuzzy/non-learning	4.33	3.96-4.71
Fuzzy/non-learning	5.43	5.14-5.72
Fuzzy/learning	8.1	7.87-8.32

Table 6