AN INTEGRATIVE FRAMEWORK FOR TIME-VARYING AFFECTIVE AGENT BEHAVIOR

Lilia Moshkina
Mobile Robot Lab, College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280, USA
lilia@cc.gatech.edu

ABSTRACT

Recent findings suggest that humans treat computers in a social way. Affect lies at the core of our social behavior; therefore, to capture this social nature of human interaction with computers and to provide a systematic way of including affect into such systems, we propose *TAME* - an integrative framework of affective agent behavior. This paper describes psychological and mathematical foundations for each of the TAME components: personality Traits, Attitudes, Moods and Emotions. An experiment to assess effectiveness of the proposed framework is also presented.

KEY WORDS

Affective modeling, behavior-based autonomous systems.

1 Introduction

Over the past few decades, computers have steadily increased their share in our lives, having come a long way from sizable mainframes accessible by a select few to indispensable artefacts used every day by the masses. With such a proliferation of computers, especially with emergence of artificial intelligent agents and personal robots in home and work environments, our attitude towards such machines has changed as well. No longer do we treat them as lifeless machines; rather, whether we like to acknowledge it or not, we treat them as social actors, towards which rules of social behavior can be applied. This proposition has been well established by Reeves [1]; in particular, people are polite to computers, are responsive (positively) to flattery and (negatively) to criticism, easily pick up various personality cues, even minimal, and accept computers as teammates [1].

Affect lies at the core of our social behavior. It should be noted that in humans the affective space is not limited to emotions only. Indeed, we often describe others in personality terms (bold or shy, a perfectionist or a slacker), refer to them as being in a "happy" or "foul" mood, and pay attention to their attitude towards us. All these phenomena, namely personality traits, attitudes, moods and emotions play an important role in our everyday behavior and comprise our affective space.

In order to capture the social nature of human interaction with computers in general, and intelligent autonomous agents in particular, and to provide a systematic way of including affect into such systems, we propose TAME - an integrative framework of affective agent behavior. TAME stands for *Traits* (personality), *Attitudes, Moods* and *Emotions* - the four components within affective space responsible for producing affective behavior. In the remainder of the paper, first the related work will be described, then the psychological and mathematical foundations for the TAME framework will be presented, followed by descriptions of an exploratory study and a proposed human-robot experiment.

2 Related Work

The official beginning of the affective computing era can perhaps be traced to Picard's[2] influential book "Affective Computing", where she advocates the usefulness of affect in machines. Since then, there's been a proliferation of work on one or another aspect of affect with the goal of improving human-computer interaction. However, only a few systems include multiple affective phenomena. The most relevant to our research system is perhaps the one proposed by El Jed Mehdi [3], as it combines emotions, personality and moods, and specifies interactions between them. The model consists of OCC-based cognitively generated emotions, Five-Factor Model based personality traits, and mood based on the current emotion and personality of the agent.

Prendinger [4] presents the SCREAM system – a scripting tool that enables authors to create emotionally and socially appropriate responses for animated characters. In this system a complex process for emotion generation (based on the OCC model), resolution, maintenance and regulation is employed which is heavily influenced by a character's personality and attitudes.

In the area of robotics, Breazeal's robotic creature Kismet [5] can be considered the first and most notable attempt at creating a socially interactive robot. In Kismet, emotions are used to enhance its interactions with people by making them more natural. Another robotic system that uses affect is Waseda Eye No. 4 Refined [6]. In this robot, a combination of emotions, moods and personality is used to provide emotional expressions for the robot's face, neck, waist and arms.

Other affective systems for embodied agents and robots include the "Oz project" [7], "Affective Reasoner" [8], "Mental Commit" Robot Paro [9], and others.

What separates the TAME framework from other work in the area is the inclusion of the entire affective space, modeling the interactions between its components and the effect of each component on behavior, and its applicability in robotics and embodied agents domains.

3 Psychological and Mathematical Foundations

Despite an increased interest among psychologists to the "feeling" part of our lives, there as yet exists no single unified theory of affect-related phenomena. Nonetheless, the existing work in the areas of Emotion, Personality, Attitude Psychology and Cognitive Science has a lot to contribute and can serve as an inspiration for endowing autonomous agents with affective capabilities, thus improving their performance in the area of human-computer interaction. The proposed framework has been initially applied to the domain of behavior-based robotics [10], but can be used more generally as an addition to any behavior-based autonomous agent architecture.

The affective module of TAME framework is composed of four interrelated components: personality Traits, affective Attitudes, Moods, and Emotions. The input into this architectural module consists of relevant perceptual information, such as the categories of visible objects and distances to them (stimuli and their strengths), as well as some internal state information. Each component (with the exception of traits that are defined a priori), runs as a separate thread continuously throughout the execution. In a behavior-based system [11], the control program consists of a collection of behaviors and coordination mechanisms; the primitive behaviors are characterized by defining parameters, and can be combined into behavioural assemblages. Instead of directly defining behavioral transitions, the affective module modifies the underlying behavioral parameters, which, in turn, directly affect currently active behaviors. The conceptual view of the framework is presented in Figure 1.

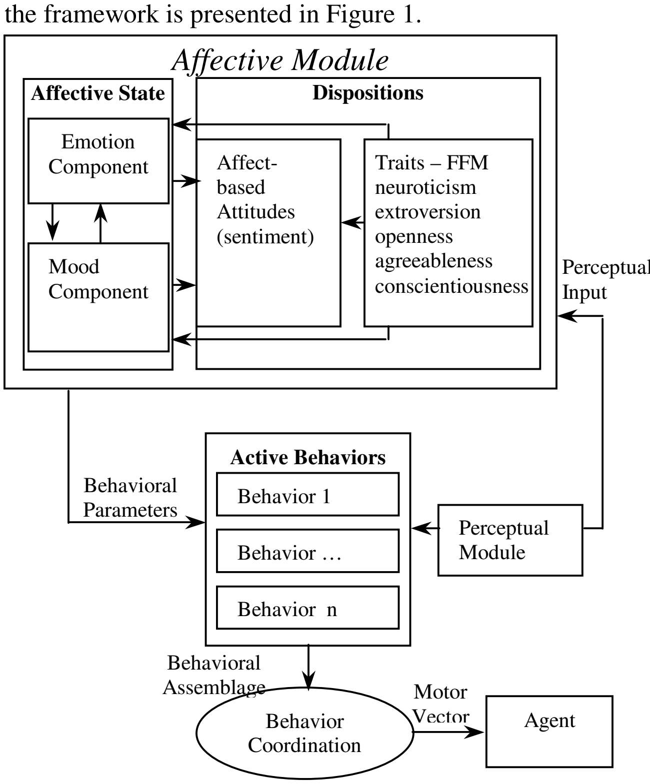

Figure 1: Integrative Framework for Affective Behavior (TAME)

Emotions and moods constitute an agent's dynamically changing, transient affective state (object-specific and short-term for emotions, and diffuse and prolonged for moods). Moods provide an affective background, or "emotional color", and can vary cyclically, whereas emotions can be viewed as "phasic perturbations on this background activity" [12]. In contrast, personality traits and attitudes are more or less time-invariant, and define general dispositions to behave and process information in certain ways. Similarly to emotions, affective attitudes, or sentiments, are object-specific; however, unlike emotions, they refer to ways of seeing and treating an object rather than to momentary responses [13]. Finally, personality refers to enduring individual differences in behavior and information processing of more general, object-independent kind. Therefore, we can position each component in the two-dimensional space defined by duration and specificity [14, 15],

Figure 2. Traits and emotions are at the opposite ends of this space: traits are time-invariant and global (independent of specific objects/events), whereas emotions are short-term, dynamically changing and focused. In addition to occupying a different position in the duration/specificity space, each of the four components also performs a distinct adaptive role: traits serve as an adaptation mechanism to specialized tasks and environments, emotions mobilize the organism to provide a fast response to significant environmental stimuli and also serve the communicative and affiliative functions, moods bias behavior according to favorable/unfavorable environmental conditions, and attitudes guide behavior towards desirable goals and away from aversive objects.

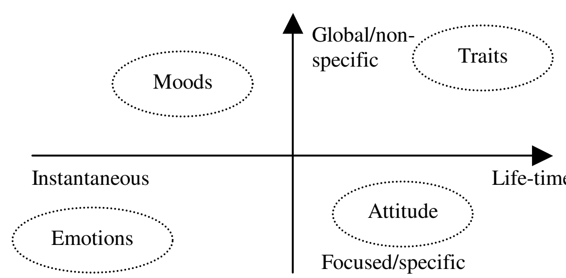

Figure 2: Relative Position of Types of Affect

3.1 Personality Traits

The Big Five taxonomy of personality traits serves as the basis for the Trait component. Personality traits refer to coherent patterns of behavior that characterize different individuals. The taxonomy is consistent over time, age and cultural differences [16], as well as applicable to nonhuman animals [17]. The five broad dimensions of personality used in the framework are represented as a vector of intensities, where intensity refers to the extent to which each trait is represented in an agent:

$$\vec{p} = [Openness, Conscientiousness, Exrtaversion, \\ Agreeableness, Neuroticism]$$

Each trait may have a direct, inverse, or no influence on agent behavior. For example, the trait of Neuroticism is directly related to avoidance behavior, whereas a trait of Conscientiousness is related to such a behavioral tendency inversely. Once the traits are defined by the user, they are then mapped to corresponding behavioral parameters according to a functional mapping $f_{ij}(p_j)$, where $B_{i,trait}$ is the new behavioral parameter, and $f_{ij}(p_j)$ is the functional mapping from the trait j to the behavioral parameter i. Figure 3 presents a second degree polynomial mapping from the trait of Neuroticism (the values are taken from the normally distributed human psychological data) onto two behavioural parameters: directly to obstacle avoidance gain (a degree to which an agent should avoid obstacles) and inversely to wander gain.

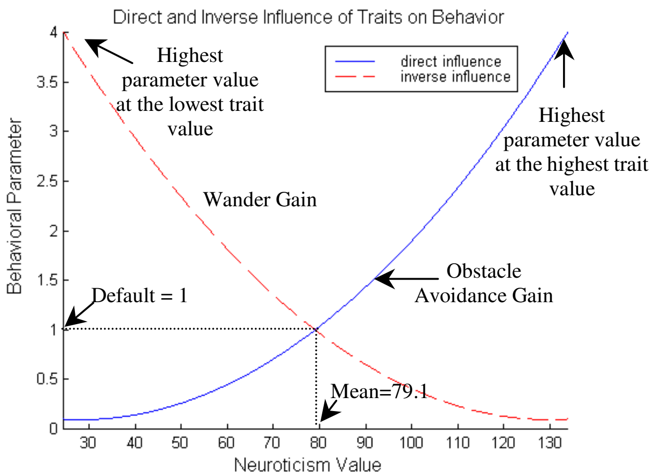

Figure 3: Comparison of Direct and Inverse Influences of Traits on Behavior

Finally, each behavioral parameter may be affected by multiple traits. In such a case, first the trait/behavior mapping for each of the influencing traits is calculated according to the chosen function $f_{ij}(p_j)$, where trait i influences behavior j, e.g., a polynomial. Then, the results are averaged across all influencing personality traits to produce the final parameter value. As the traits are relatively time-invariant, the obtained trait-based behavior parameters serve as default behaviors for the agent.

3.2 Emotions

Emotion can be described as a coherent and organized mechanism or system that, under environmental contingencies, provides an organism with a fast, flexible, adaptive response by guiding, organizing, or coordinating behavior, perception, and communication. Picard [2] identifies a number of properties of emotions that would be desirable to model in affective systems; the following subset of them is modeled in TAME:

1. **A property of activation.** Refers to certain stimulus strength below which the emotion is not activated.
2. **A property of saturation**. Refers to the upper bound of an emotion, after which, regardless of the increasing stimulus strength, the emotion doesn't rise any more.
3. **A property of response decay**. States that emotions decay naturally over time unless they are re-stimulated.
4. **A property of linearity**. Emotions can be modeled as linear under certain conditions; due to the properties of activation and saturation, the emotions will approximate linearity only for certain stimulus strength range, and will approach a sigmoid at its edges.
5. **Personality and Mood Influences**. Personality may influence the threshold of eliciting stimulus (activation point), peak (amplitude) of response, and rise time to peak [18]. Moods can vary the threshold for experiencing a particular emotion [13].

The basic emotions used in the framework are *fear, disgust, anger, sadness, joy and interest*. Each emotion's intensities are stored in the emotion intensity matrix:

$$\vec{E} = [E_i]$$

where $0 \le E_i \le g_i$, the value E_i represents the intensity of a currently active emotion, 0 signifies the absence of emotion, and g_i is the upper bound for emotion i.

The eliciting stimulus strength for each emotion is calculated by taking into account a number of object properties, such as its physical properties (size, shape, etc.), its position (distance to the object, its velocity, etc.), and any existing attitude of the agent towards the eliciting stimulus. Based on the current stimulus strength and personality and mood influences, the base emotion level is calculated as follows:

$$E_{i,base} = \begin{cases} e^{(s_i - a_i)/d} - e^{(-a_i)/d_i}, & if \quad a_i \le s_i < (a_i + b_i)/2 \\ g_i - e^{(s_i - b_i)/d_i}, & if \quad s_i \ge (a_i + b_i)/2 \end{cases}$$

$$where \quad b_i = 2 \cdot d_i \cdot \ln(g_i + e^{(-a_i)/d_i})/2 + a_i$$

where $E_{i,base}$ is the base emotion value for emotion i, s_i is the strength of stimulus eliciting emotion i, a_i is the variable that controls the activation point for emotion i, d_i is the variable that controls the maximum slope for emotion i, g is the amplitude of emotion i, and b_i is the break-point, at which the emotion reverses its rate of growth. Figure 4 presents the resulting curve graphically.

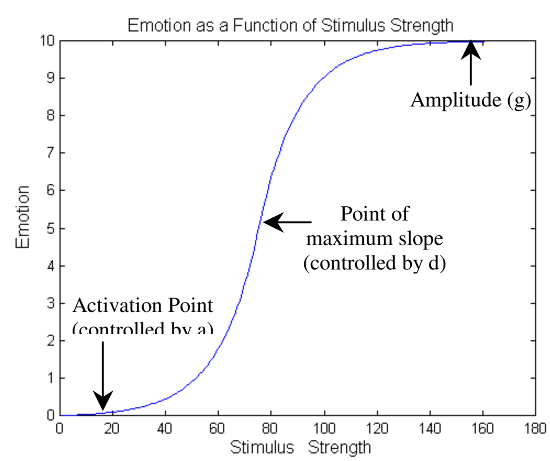

Figure 4: Emotion Generation Based on Stimulus Strength

A linear mapping from traits to amplitude, activation point, and maximum slope is used to obtain personality influence on emotion generation. For example, the trait of Extraversion provides a direct influence on the amplitude, activation point, and slope of positive emotions (joy and interest), therefore an agent with a higher level of Extraversion will have a stronger positive emotion that will be activated at a weaker stimulus strength and will rise faster than that of an introverted agent. Similarly, current mood will influence the activation point, where the negative mood will make it easier for an agent to experience negative emotions, and positive mood – positive emotions. Figure 5 presents combined influence of traits and mood on emotion generation.

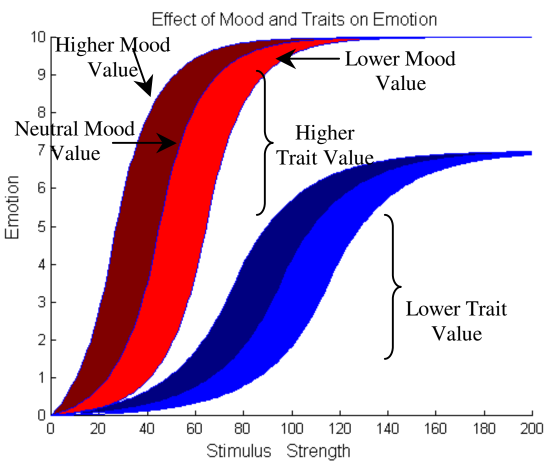

Figure 5: Combined Mood and Trait Influences on Emotion Generation

To account for the short-term duration of emotions and habituation to prolonged stimulus, emotion decay is modeled as a slowly decreasing exponential:

$$E_{i,t,decay} = E_{t,base} - e^{(t-t_0)*d}$$

where $E_{i,t,decay}$ is the intensity of emotion i at time t, t_o is the time at which emotion is activated (becomes greater than 0), and d is a variable that controls the rate of decay.

Finally, in order to smooth the emotion change in cases of sudden appearance and disappearance of eliciting stimuli, a weighted averaging filter can be used:

$$E_{i,t,filtered} = (w_{current} * E_{i,t,decay} + w_{prior} * E_{i,t-1,filtered}) / (w_{current} + w_{prior})$$

where $E_{i,t,filtered}$ is the final intensity of emotion i at time t after filtering, $w_{current}$ and w_{prior} are weighting variables controlling the relative importance of current and previous emotional states. This filtering function will help to account for short-term lingering emotions even after the eliciting stimulus has disappeared.

3.3 Moods

As opposed to emotion, which is a short-term, high activation state that appears as a result of a stimulus, mood can be viewed a continuous low activation variable affective state, or "stream of affect", and can be characterized as diffuse or global [Frijda #51]. According to Watson [15], mood can be represented along two different dimensions, Positive Affect and Negative Affect, where Negative Affect refers to the extent to which an individual is presently upset or distressed, and Positive Affect generally refers to one's current level of pleasure and enthusiasm. The level of arousal for both categories can vary from low to high; a low positive mood value has a negative connotation ("sluggish", "disinterested") and refers to insufficient level of pleasure and enthusiasm, rather than just low. Conversely, as negative affect signifies how upset or distressed an individual is, low level of negative mood would suggest that the individual is calm, rather than distressed.

There are two broad types of mood change: environmental - external (level of light, noise, etc.) and internal (e.g., battery level in case of a robot); and short-term situational variables, including emotional episodes. The current base level of mood of an agent is defined as a weighted summation of various external and internal variables. Assuming that the same variables affect both positive and negative moods, strengths of environmental and internal inf*l*uences can be represented in a matrix $\vec{l} = [l_i]$, where $0 < l_i < b_i$, where b_i is the hardware-dependent upper bound (e.g., light can only be detected up to a certain level, etc.). The relative weights for each variable are stored in the mood generation matrix $\overrightarrow{mg} = [mg_{ij}]$. The values in this matrix are unit conversion factors, to convert the various mood generation variables to the same unit, and should be found experimentally for each variable. In addition, negative mg_i stands for inverse influence of the variable on the mood, and positive mg_i stands for direct influence. As moods are continuous, always present streams of affect, the base mood can be continuously generated based on the current environmental and internal influences as follows:

$$m_{base} = \begin{bmatrix} m_{positive} \\ m_{negative} \end{bmatrix} = \sum_{i=1}^{N} \begin{bmatrix} mg_{i,positive} \\ mg_{i,negative} \end{bmatrix} \cdot (l_i - l_{i,neutral})$$

where $\overrightarrow{mg}$ is the mood generation matrix, $\vec{l}$ is the mood generation variable strength matrix, and N is number of mood generation variables. Figure 6 illustrates the effect of an environmental variable on mood.

Additionally, similarly valenced emotions can affect the corresponding mood intensities additively, and change the existing base mood level in the following manner:

$$m_{positive} = m_{positive\,base} + E_k, \ \ if\ E_k \in \{Interest, Joy\}$$
$$m_{negative} = m_{negative\,base} + E_k, \ \ if\ E_k \in \{Fear, Anger, Disgust, Sadness\}$$

where $m_{positive}$ is the emotion-based intensity of positive mood, $m_{positive}$ is the emotion-based intensity of positive mood, and $\vec{E}$ is the emotion intensity matrix.

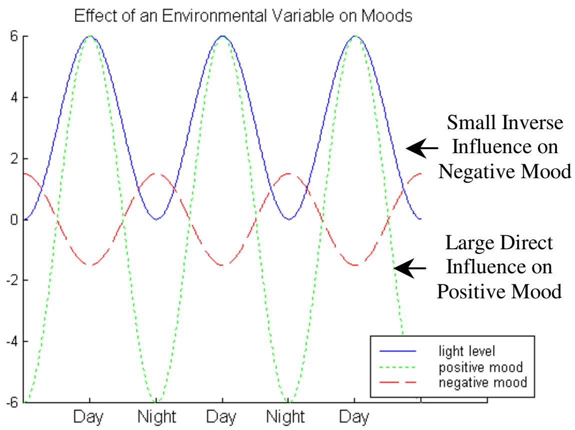

Figure 6: **Different Effect of Light on Positive and Negative Mood**

Finally, as mood is a low-activation, slow-varying affective state, a filter is used to smooth the effect of emotions and environmental and internal influences.

3.4 Affective Attitudes

Affective attitude can be defined as a "learned predisposition to respond in a consistently favorable or unfavorable manner with respect to a given object" [19]. Such an attitude can be represented as a single value *A*, ranging from $-\infty$ to ∞, where 0 represents a neutral (or absence of) attitude, negative values represent increasingly strong negatively valenced attitude (ranging from a mild dislike to hatred), and positive values refer to increasingly strong positively valenced attitude (e.g., from a subtle like to adoration). Attitudes are object-specific, and an initial attitude for a particular object (*y*) would consist of a combination of positive or negative attributes of this object, represented as a matrix $\vec{o}_y = [o_{iy}]$, where $-\infty < o_{iy} < \infty$. Such attributes are not limited to properties of the object only; for example, an emotion invoked by the object and any actions taken by the object may be considered "attributes". The initial value of the attitude for object *y* ($A_{y,init}$) is calculated as follows:

$$A_{y,init} = \sum_{i=1}^{N} o_{iy}$$

where $A_{y,init}$ is the newly-formed attitude for object *y*, o_{iy} is an attribute *i* of object *y* that is involved in the attitude formation, and *N* is the number of attributes for object *y*.

Assuming that initial impression is the strongest, substantial changes in attitude are fairly hard to achieve, therefore any subsequent exposure to the same object would result only in incremental change. This is achieved through discounting any additional positive or negative object attributes to a certain extent. The updated attitude value for object *y* for n-th encounter ($A_{y,n}$) would then be calculated as follows:

$$A_{y,n} = A_{y,n-1} + \lambda_n \cdot (\sum_{i=1}^{N} o_{iy})$$

where $A_{y,n-1}$ is the attitude towards object *y* at encounter *n-1*, *n* is the total number of encounters up to date, $\vec{o}_y$ is the matrix of attributes for object *y*, and λ is the discount factor.

Consistent with the finding on mood-congruent judgment, positive mood increases the value of the attitude (a_y) towards an object *y*, and negative mood – decreases it as follows:

$$A_{y,mood} = A_y + K(m_{positive} - m_{negative})$$

where $A_{y,mood}$ is mood-enhanced value of agent's attitude towards object *y*, A_y is the original value of agent's attitude towards object *y*, $m_{positive}$ is the current positive mood value, $m_{negative}$ is the current negative mood value, and *K* is a scaling factor to bring moods and attitudes to the same units.

4 Exploratory Experimental Study

In order to assess human attitude towards affective systems, and therefore the potential usefulness of the TAME framework, an exploratory longitudinal study was conducted (described in detail in [20]). In this study, 20 participants interacted with a Sony robotic dog Aibo ERS-210 in 4 sessions, to allow the participants to "bond" with the robot. The study had two between-subject conditions: presence of affect (partially implemented trait and emotion components were used to encode affect) and absence of affect; 10 males and 10 females were distributed equally between the conditions.

The participants' impressions were measured by a means of the post questionnaire ([20]), and positive/negative emotionality questionnaire to assess mood [21]. Although there were no statistically significant differences as to whether or not the participants could perceive personality or emotion in the robot, they nonetheless found it easier to interact with Aibo in the affective condition ($p<0.004$), as well as their negative mood was lower ($p<0.048$) in the affective condition, signifying lower levels of distress and frustration. Also, those who thought that the robot expressed affect attributed their enjoyment of interaction partially to this feature. Interestingly, women were significantly more prone to attribute emotions to the robot than men, regardless of the condition ($p<0.043$).

5 Proposed Human-Robot Interaction Experiment

The results of the initial study were encouraging as to potential benefits of affect in human-robot interaction; therefore, a larger human-robot interaction experiment has been designed. For this experiment, the framework will be fully integrated into the *MissionLab* system, which is a version of *AuRA* (Autonomous Robot Architecture) [22], and implemented on a Pioneer-2 DXE robot. The experiment will contain three within-subject conditions: No Affect, Emotions Only, and Emotions, Personality and

Attitudes Combined. In addition to these three within-subject conditions, the Extraversion dimension of personality will be examined as a separate between-subject factor as applied to the multi-affective condition of the within-subject portion of the experiment.

A cooperative foraging task will be used for this experiment, in which participants will be asked to help the robot locate and collect foraging objects. In the affective conditions, the robot will exhibit the emotions of fear (at a loud sound next to a large object), interest (exploring the environment), and joy (at finding the foraging objects). The extraverted personality will be expressed through smaller distance between the robot and the human, and more pronounced expressions of positive emotions.

With the help of this experiment, we hope to uncover whether complex affective behavior provides any benefits beyond a simpler emotions-only behavior, especially in terms of ease of use, pleasantness of interaction, and acceptance of robotic technology by novices.

6 Conclusion

In this paper, we presented an integrative framework of affective agent behavior, TAME. The framework combines a wide variety of affective phenomena (traits, attitudes, moods and emotions), is psychologically inspired, models interactions between the phenomena and their effect on behavior, and can be applicable to a variety of domains, including robotics and intelligent agents. In order to assess the advantages of the TAME framework, a human-robot experiment has been designed and will be conducted in the near future.

References

1. Reeves, B., Nass, C., *The Media Equation* (1996, Cambridge: Cambridge University Press).
2. Picard, R.W., *Affective Computing* (1997, The MIT Press: Cambridge, MA).
3. El Jed Mehdi; Pallamin, N., Dugdale, J., Parvard, B., Modelling Character Emotion in an Interactive Virtual Environment. *AISB 2004 Convention, Symposium on Language, Speech and Gesture for Expressive Characters, 2004*.
4. Prendinger, H., Descamps, S., Ishizuka, M., Scripting affective communication with life-like characters in web-based interaction systems. *Applied Artificial Intelligence, 16(7-8)*, 1996, p. 519-553.
5. Breazeal, C., Emotion and Sociable Humanoid Robots. *International Journal of Human-Computer Studies, 59*, 2003, p. 119-155.
6. Miwa, H. et al., Effective Emotional Expressions with Emotion Expression Humanoid Robot WE-4RII. *IROS, 2004*.
7. Bates, J., The Role of Emotion in Believable Agents. *Communications of the ACM, Special Issue on Agents, 37(7)*, 1994, p. 122-125.
8. Elliott, C., Siegle, G., Variables Influencing the Intensity of Simulated Affective States. *AAAI Spring Symposium on Reasoning about Mental States: Formal Theories and Applications,* 1993: p. 58-67.
9. Wada, K., Shibata, T., Saito, T., Tanie, K., Effects of Robot-Assisted Activity for Elderly People and Nurses at a Day Service Center. *Proceedings of the IEEE*, *92(11),* 2004, p. 1780-1788.
10. Moshkina, L., Arkin, R.C., On TAMEing Robots. *Interanational Conference on Systems, Man and Cybernetics*, 2003.
11. Arkin, R., Motor-Schema based Mobile Robot *Navigation. International Journal of Robotics Research, 8(4),* 1989, p. 92-112.
12. Davidson, R.J., On Emotion, Mood and Related Affective Constructs, in *Nature of Emotion: Fundamental Questions* (D. Ekman P., R., Editor. 1994, Oxford Unviversity Press: New York).
13. Frijda, N.H., Varieties of Affect: Emotions and Episodes, Moods, and Sentiments, in *The Nature of Emotion: Fundamental Questions* (P. Ekman, Davidson, R.J., Editor. 1994, Oxford University Press).
14. Moffat, D., Personality Parameters and Programs. *Creating Personalities for Synthetic Actors*, 1997: p. 120-165.
15. Watson, D., Clark, L.A., Tellegen, A., *Mood and Temperament* (2000: The Guilford Press).
16. McCrae, R.R., Costa, P.T., Toward a new generation of personality theories: theoretical contexts for the five-factor model. *Five-Factor Model of Personality* (Wiggins, J.S., Editor. 1996, The Guilford Press).
17. Gosling, S.D., John, O.P., Personality Dimensions in Nonhuman Animals: A Cross-Species Review. *Current Directions in Psychological Science*, 1999: p. 69-75.
18. Davidson, R.J., Affective Styles and Their Cognitive Consequences, in Handbook of Cognition and Emotion (T. Dalgleish, Power, M.J., Editor. 1999, John Wiley and Sons. p. 103-124).
19. Breckler, S.J., Wiggins, E.C., On Defining Attitude and Attitude Theory: Once More with Feeling, in *Attitude Structure and Function* (A.R. Pratkanis, Breckler, S.J., Greenwald, A.G., Editor. 1989, Lawrence Erlbaum Associates. p. 407-429).
20. Moshkina, L., Arkin, R.C. Human Perspective on Affective Robotic Behavior: A Longitudinal Study. *Proc IEEE IROS*. 2005. Edmonton, Canada
21. Watson, D., Clark, L.A., Tellegen, A., Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality & Social Psychology, 56(6)*, 1998, p. 1063-1070.
22. Arkin, R., Balch, T., *AuRA:* Principles and Practice in Review. *Journal of Experimental and Theoretical Artificial Intelligence*, 9(2), 1997, p. 175-189.