Modeling Emotions and Other Motivations in Synthetic Agents

Juan D. Velásquez
MIT Artificial Intelligence Laboratory
545 Technology Square, NE43-812, Cambridge, MA 02139
jvelas@ai.mit.edu

Abstract

We present *Cathexis*, a distributed, computational model which offers an alternative approach to model the dynamic nature of different affective phenomena, such as emotions, moods and temperaments, and provides a flexible way of modeling their influence on the behavior of synthetic autonomous agents. The model has been implemented as part of an extensible, object-oriented framework which provides enough functionality for agent developers to design emotional agents that can be used in a variety of applications including entertainment (e.g. synthetic agents for interactive drama, video games, etc.), education (e.g. Intelligent Tutoring Systems), and human-computer interfaces.

Introduction

Emotions are an essential part of our lives, they influence how we think and behave, and how we communicate with others. Several researchers have acknowledged their importance in human thinking [Minsky 1986; Toda 1993], and recent neurological evidence seems to support these ideas [LeDoux 1996; Damasio 1994].

Recent developments in the areas of synthetic agents, [Maes 1995; Blumberg 1994; Bates 1994; Elliot 1992; Reilly 1996;] and affective computing [Picard 1995], have promoted the study of emotions and their influences in behavior and learning [Velásquez 1996; Kitano 1995]. Nevertheless, up to date, relatively few computational models of emotion have been proposed. For a review of some of these models see [Pfeifer 1988].

This paper describes a computational model that focuses on different aspects of the generation of emotions and their influence on the behavior of synthetic agents.

The Cathexis Model

We have developed a distributed model for the generation of emotions and their influence in the behavior of autonomous agents. The model is called Cathexis (concentration of emotional energy on an object or idea) and has been inspired by work in different fields including among others, Psychology, Ethology and Neurobiology. Although its main objective is to model several aspects of the generation of emotions, the model also provides simple models for other motivations and an algorithm for action-selection.

Figure 1 provides a high level view of the model's architecture.

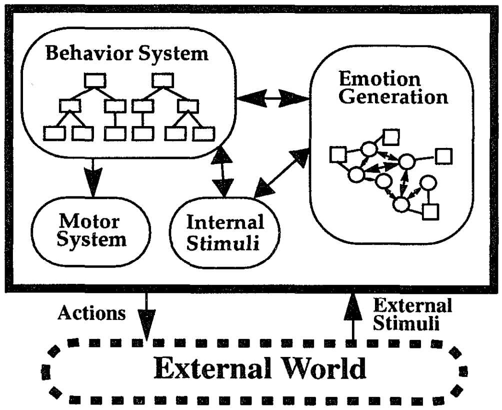

Figure 1 The Cathexis Architecture.

Modeling Affect: The Emotion Generation System

Emotions, moods, and temperaments are modeled in Cathexis as a network composed of special emotional systems comparable to Minsky's "proto-specialist" agents [Minsky 1986]. Each of these proto-specialists represents a specific emotion family (See below) such as *Fear* or *Disgust*. Within each proto-specialist, different sensors are monitoring both external (e.g. events in the environment) and internal (e.g. drive levels, feedback from sensorimotor processes) stimuli for the existence of the appropriate conditions that would elicit the emotion represented by that particular proto-specialist. These sensors are arranged into four different groups (*Neural*, *Sensorimotor*, *Motivational*, and *Cognitive*) that represent different kinds of both cognitive and noncognitive emotion activation systems as it will be described below. Input from these sensors either increases or decreases the intensity of the emotion proto-specialist to which they belong. Associated with each proto-specialist are two threshold values. The first threshold, α, controls the activation of the emotion. That is, once the intensity goes above this threshold, the emotion proto-specialist releases its output signal to other emotion proto-specialists, and to the Behavior System which in turn selects an appropriate behavior according to the state of these emotional systems.

The second threshold, ω, specifies the level of saturation for that emotion proto-specialist. This is consistent with real life emotional systems in which levels of arousal will not exceed certain limits. Another important element associated with an emotion proto-specialist is a decay function, $\Psi()$, which controls the duration of the emotion episode as it will be explained below. Figure 2 illustrates some of these ideas.

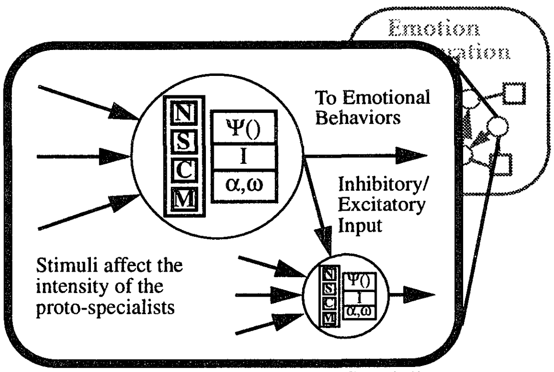

Figure 2 Emotion Proto-Specialists.

All of these emotion proto-specialists run in parallel and are constantly updating their intensities. No particular proto-specialist is in control of the system. In fact, more than one may be active (i.e. have intensity values above zero) at the same time. Nonetheless, when their intensity value goes above the α threshold they may inhibit or excite other proto-specialists. This type of interaction is consistent with real life emotional systems, in which high arousals in one or more systems will tend to inhibit or excite other emotional systems (e.g Fear inhibits Happiness, Happiness and Sadness inhibit each other, etc.).

Basic Emotions. Addressing the issue of what emotions to include in the model we have drawn on ideas from Ekman, Izard, and several other theorists [Ekman 1992; Izard 1991; Johnson-Laird and Oatley 1992], and have provided explicit models for the so-called *basic* or *primary* emotions. The expression *basic emotions* has been used by researchers in many different ways. In this model the term basic, as in [Ekman 1992], is used to claim that there are a number of separate discrete emotions which differ from one another in important ways, and which have evolved to prepare us to deal with fundamental life tasks, such as falling in love, fighting, avoiding danger, adjusting to losses, and so on.

Within Cathexis, each basic emotion is not a single affective state but rather a family of related affective states (e.g. Fear, Fright, Terror, Panic, etc.). Each member of an emotion family shares certain characteristics, such as similarities in antecedent events, expression, likely behavioral response, physiological activity, etc. These characteristics differ between emotion families, distinguishing one from another. Cathexis includes an emotion proto-specialist for each of the following basic emotions: *Anger*, *Fear*, *Distress/ Sadness*, *Enjoyment/Happiness*, *Disgust*, and *Surprise*.

Emotion Blends and Mixed Emotions. Certainly, some affective states considered by many to be emotions are not included in the list defined above. The concept of emotion families would include some of these states as variations (in intensity, expression, etc.) of a basic emotion, but some other affective states do not seem to fit within these categories. Although some theorists claim that all emotions are basic, there is some evidence that suggests the existence of *blends*, a term which describes the co-occurrence of two or more basic emotions at a time. Examples of blends might be emotions such as *Grief* and *Jealousy*. *Grief*, for example, is certainly more specific than *Sadness*. Besides a variation of *Sadness*, people who experience *Grief* are also likely to feel *Anger* and *Fear*, and perhaps even *Surprise* depending on the situation.

As mentioned before, more than one emotion proto-specialist can be active at the same time, which means that two or more basic emotions may co-occur, representing, as a whole, emotion *blends* and *mixed* emotions. The intensity level and the influences (both in expression and in experience) of each of the active emotions, give rise to these secondary emotions.

Cognitive and Noncognitive Elicitors of Emotion. Several models or taxonomies for analyzing the conditions leading to emotions have been proposed. Appraisal theorists, such as [Frijda 1986; Ortony, Clore, & Collins 1998; Roseman, Spindel, and Jose 1990], claim that what determines whether an emotion will be felt and what emotion it will be, depends on the evaluation and interpretation of events, rather than on the events per se. Although only a few of these authors take an extreme position in which cognition is the exclusive determinant of the emotions, most theories on the generation of emotion are concerned only with cognitive elicitors.

In contrast, and with a strong influence from Izard's multi-system for emotion activation [Izard 1993], and also drawing on ideas from Tomkins [Tomkins 1962], Cathexis takes into account the existence of both cognitive and noncognitive elicitors of emotions. These elicitors are arranged into four categories as described by Izard:

- *Neural*: Includes the effects of neurotransmitters, brain temperature, and other neuroactive agents that can lead to emotion and which can be mediated by hormones, sleep, diet, environmental conditions, etc.
- *Sensorimotor*: This system covers sensorimotor processes, such as facial expressions, body posture, muscle action potentials, and central efferent activity, that not only regulate ongoing emotion experiences but can also elicit emotion.
- *Motivational*: This system includes all motivations that lead to emotion. In this model, motivations include drives (e.g. *Thirst* and *Hunger*), emotions (e.g. *Anger*, and *Happiness*), and pain regulation. Some examples of elicitors in this system include the innate response to foul odors or tastes producing disgust, pain or aversive stimulation causing anger, and emotions like sadness eliciting others like anger, and vice versa.
- *Cognitive*: This system includes all type of cogni-

tions that activate emotion, such as appraisal and interpretation of events, comparisons, attributions, beliefs and desires, memory, and so on. In Cathexis, cognitive elicitors for the different basic emotions are based on an adapted version of Roseman's revised theory [Roseman, Spindel, and Jose 1990]. Several reasons influenced the decision for choosing Roseman's theory and not one of the others. First, Roseman's theory is broad enough in scope. That is, it attempts to address all emotion-eliciting stimuli that people normally face. Second, it presents clear testable predictions for the relationships between specific appraisal configurations and the emotional states they produce. Furthermore, these predictions are supported by empirical research as can be seen in [Roseman, Spindel, and Jose 1990]. And third, it fits well within the multi-system activation model described here since Roseman assumes that all emotions have a motivational basis, and that in some cases these motivations may be noncognitive.

By considering both cognitive and noncognitive elicitors of emotion, the proposed model is more robust and gives developers more flexibility in defining the agent's affective characteristics.

Moods and Other Affective Phenomena. A number of words considered to be emotion terms actually refer to moods rather than emotions (e.g. nervousness, irritation, euphoria). Sometimes, even the words emotion and mood are used interchangeably by some psychologists and laypeople alike, to refer to certain aspects of affect. However, moods and emotions interact in important ways and may differ in several aspects including their function.

Following a psychobiological perspective, Cathexis differentiates between moods and emotions in terms of levels of arousal. Emotions may consist of high arousal of specific types of brain systems. These systems, however, may be tonically activated at low levels by a variety of internal stimuli, as well as relatively weak perceptual inputs. Thus, moods may be explained as low tonic levels of arousal within emotional systems, which as we have mentioned are modeled in Cathexis as emotion proto-specialists, while emotions would be explained as high levels of arousal in these same systems [Panksepp 1994].

Thus, while high arousal of emotion proto-specialists will tend to inhibit other proto-specialists, mild arousal may very well allow several systems to be concurrently active, leading to the chance of an enormous display of mood states compared to the limited number of basic emotional states. This representation is consistent with the enormous subtleties of human moods and feelings. It is also consistent with the common observation that moods seem to lower the threshold for arousing certain emotions because emotional systems that are aroused, as it happens in the representation of moods, are already providing some potential for the activation of an emotion. Finally, it is consistent with the observation that the duration of moods appears to be longer than that of the emotions, since at low levels of arousal, the intensity of the emotion proto-specialists will decay more slowly.

Temperaments are modeled through the different values the activation and saturation thresholds can have for the different proto-specialists. Thus, for instance, an individual who has propensities to be in fearful moods might have a lower threshold for the emotion of fear in comparison to other individuals who do not.

Emotion Intensity. We mentioned the idea of thresholds and levels of arousal. But what determines the intensity of an emotion? This question has hardly been a subject of research. Besides the analyses by Ortony, Clore, and Collins [Ortony, Clore, and Collins 1988], Frijda [Frijda 1996], and Bower [Bower and Cohen 1982], emotion intensity as such, has not received a lot of attention.

Understanding how the intensity of an emotion is affected amounts to specifying the precise nature of each of the factors that contribute to it. Thus, in Cathexis, the intensity of an emotion is affected by several factors, including the previous level of arousal for that emotion (which takes into account the mood), the contributions of each of the emotion elicitors for that particular emotion, and the interaction with other emotions (inhibitory and excitatory inputs). All of this is modeled as described in equation (1):

$$I_{et} = \chi\left(\Psi(I_{et-1}) + \sum_k L_{ke} + \sum_l G_{le} \cdot I_{lt} - \sum_m H_{me} \cdot I_{mt}\right) \quad (1)$$

Where I_{et} is the value of the intensity for Emotion e at time t; I_{et-1} is its value at the previous time step; $\Psi()$ is the function that represents how Emotion e decays; L_{ke} is the value of Emotion Elicitor k (See [Velásquez 1996] for more details on how these are computed), where k ranges over the Emotion Elicitors for Emotion e; G_{le} is the Excitatory Gain that Emotion l applies to Emotion e; I_{lt} is the intensity of Emotion l at time t; H_{me} is the Inhibitory Gain that Emotion m applies to Emotion e; I_{mt} is the intensity of Emotion m at time t; and $X()$ is the function that constrains the intensity of Emotion e between 0 and its saturation value.

Emotion Decay. If the concept of emotion intensity has received little attention, the concept of emotion decay has received much less. Perhaps this is why most previous models of emotion, with the exception of that proposed by Reilly [Reilly, 1996], do not consider this issue. With the current evidence on this matter one can only speculate as to what causes emotions to decay and what are the mechanisms and the systems involved in this process.

In Cathexis, each emotion proto-specialist has been provided with a built-in decay function that controls the duration of the emotion once it has become active. This function is specific to each emotion proto-specialist which allows for different models of emotion decay. That is, the decay function for an emotion can be implemented as a constant of time or as a more complex function of the different elicitors for that particular emotion (e.g. in terms of resolution of goal-centered issues). Regardless of their implementation, in every update cycle the decay functions of the active emotions are evaluated, and unless there is some excitatory input for those emotions, their intensity is lowered and after a few cycles they become inactive.

Behavior System

Researchers such as Frijda, suggest that emotional systems may elicit physiological changes, and they may elicit feelings, but unless they imply or involve changes in the readiness of specific behaviors they are of no consequence for either the individual or for others [Frijda 1996]. These changes are modeled in the Behavior System, which decides what behavior is appropriate for the agent to display, given its emotional state at some point in time.

The Behavior system is a distributed system composed of a network of behaviors, such as "fight", "kiss", or "smile". Each of these behaviors competes for the control of the agent. The decision of which behavior should be active is based on the value of each of the behaviors, which is recalculated on every cycle as described below. Each behavior contains two major components: the Expressive or Motor Component and the Experiential Component.

Expressive Component. Certainly, the expression of emotion is one of the most important aspects in emotional behavior. In Cathexis, the expressive component of a behavior contemplates the following aspects:

- *Prototypical facial expression*: Motor commands to alter the facial expression accordingly.
- *Body posture*: Series of motor commands to alter the agent's body posture accordingly.
- *Vocal expression*: May consist of non-language sounds such as laughing, yawning, grunting, etc., or it may include more complex elements to control speech, such as loudness, pitch, and temporal variability.

Experiential Component. The experience of emotions and moods bias action and cognition. Several researchers agree that the experiential component of emotions and moods can be identified as motivational processes that manifest themselves as action readiness, perceptual biases, selective filters for memory and learning, and a feeling state [Bower and Cohen 1982; Izard 1993]. The main aspects considered in this model include:

- *Motivation*: Motivations are specified as the main causes of activation or release of behaviors. They are also specified within the experiential component because a behavior, when active, may influence different motivational systems. Behaviors affect the levels of drives, but they may also affect emotions and moods as it was discussed with sensorimotor elicitors of emotion.
- *Action tendency*: Changes in action readiness or action tendencies are a major, if not the major, aspect of the response to an event of emotional significance [Frijda 1986]. In Cathexis, action tendency and readiness are modeled by the behavior itself. Thus, a behavior, such as aggression, which may be represented as several different behaviors including perhaps fighting, insulting, biting, etc., is modeled as a direct response to the emotion of anger, or an irritable mood.

Other influences of emotions and moods, such as *perceptual biases and selective filtering* include complex interactions with cognitive systems and constitute by themselves whole fields of active research which are beyond the scope of this work. Nevertheless, we believe it is important to acknowledge their existence and we have tried to design Cathexis as an open model that allows incremental additions of systems or components that model these influences.

Selection of Behaviors. As mentioned before, behaviors compete with each other to obtain control of the agent. Competition is based on the values of each behavior. At some point in time, the behavior with the highest value becomes the active behavior. This value is updated every cycle and depends on several factors, called the *releasers* for that behavior, which may include motivations (emotions, moods, drives, pain), as well as a wide variety of external stimuli. The idea of releasers or releasing mechanisms is taken from the field of Ethology, primarily, from the work of Tinbergen [Tinbergen 1969] and it has already been successfully used in models of action-selection such as that proposed by Blumberg [Blumberg, 1994].

The value of a behavior is modeled as the sum of the values of each of all Behavior Releasers relevant to that particular behavior. The value of each of these releasers depends on the specific nature of the releaser. For instance, if the releaser is an internal motivation, its value is calculated in terms of the intensity of that motivation. On the other hand, if the releaser consists of some sort of external stimuli, then its value depends on sensory input which determines if the stimulus of interest is present and matches some specific criteria.

This model of behavior is certainly simple. A more complex model may include inhibitions between behaviors in a similar fashion to that of inhibitory inputs for emotional systems, and it may also include models of behavior depletion or fatigue as those described by Ludlow [Ludlow 1980]. It should be noted, however, that this part of the model has intentionally been made simple and general enough so that Cathexis can be integrated into several different architectures of action-selection, such as those of Maes [Maes 1991], and Blumberg [Blumberg 1994].

The Algorithm. Emotion generation and action-selection is done as described below. On each update cycle:

1. Both the internal variables (i.e motivations) and the environment are sensed.
2. The values for all of the agent's motivations (both drives and emotions) are updated as described above.
3. The values of all behaviors are updated based on the current sensory stimuli (external stimuli and internal motivations).
4. The behavior with the highest value becomes the active behavior. Its expressive component is used to modify the agent's expression, and its experiential component is

evaluated in order to update all appropriate motivations.

Simón the Toddler: A Testbed Environment

Cathexis has been implemented in its totality as part of an object-oriented framework that allows agent developers to create emotional agents. This framework has been implemented in C++ on a Power Macintosh 8500 and has been used to build an environment in which the user interacts with "Simón", a synthetic agent representing a young child. The main goal of creating Simón was not to create an emotional, believable agent, but rather to build an environment that would serve the purpose of a testbed in which we could experiment and test the internals of the Cathexis model, while at the same time, evaluate how useful is the framework as a tool to create models of emotional phenomena and emotional agents.

Simón's Motivations and Behaviors

In order to test the relationships between motivations, Simón was designed to have several instances of different motivational systems: five drive proto-specialists (*Hunger, Thirst, TemperatureRegulation, Fatigue*, and *Interest*), and six emotion proto-specialists (*Enjoyment/Happiness, Distress/Sadness, Fear, Anger, Disgust*, and *Surprise*).

Similarly, Simón has a repertoire of approximately fifteen behaviors, such as *Sleep, PlayWithToy, Eat, Drink, Laugh, Cry, Hug, Kiss*, and so on, which correspond to action tendencies in response to the state of his motivational systems. Each of these behaviors have different expressive and experiential components as described above. The expressive component includes different facial expressions similar to those shown in Figure 3, and for some behaviors it also includes vocal expressions in the form of non-language sounds, such as a cry, or a giggle. The experiential component includes influences in Simón's motivations, (e.g. modifying the intensity of particular drives), and specific actions particular to each behavior (e.g. biting, waving hand, and so on).

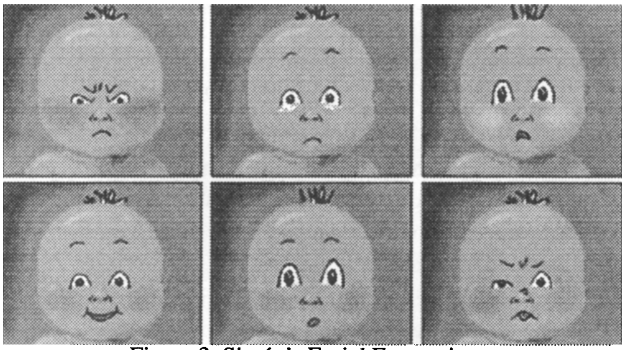

Figure 3 Simón's Facial Expressions

Interacting with Simón

The user interacts with Simón by performing actions in the environment using different controls, such as sliders, icons, buttons, and menus. These actions may vary from things such as altering the level of Simón's neurotransmitters and hormones, to higher level actions, such as feeding him or changing the lighting and temperature of his room. These interactions provide stimuli to Simón, which, along with the internal stimuli generated by his motivational systems will cause him to react emotionally. For instance, when his desire to play is high, and there are toys available for him to play with, Simón becomes happy. Similarly, when the level of *Hunger* is high and there is no food around, his level of *Distress* increases and eventually becomes sad and angry. Once this need is satisfied, his level of *Anger* and *Sadness* decays, and he becomes happy.

The user can also manipulate the different parameters within Simón's proto-specialists (See Figure 2). Distinct configurations of these elements allow for the representation of different emotional reactions, moods, and temperaments, all of which we refer to as the *affective style* of the agent. Thus, playing with these values may give the user the impression that Simón is suffering from a severe case of multiple personality disorder!

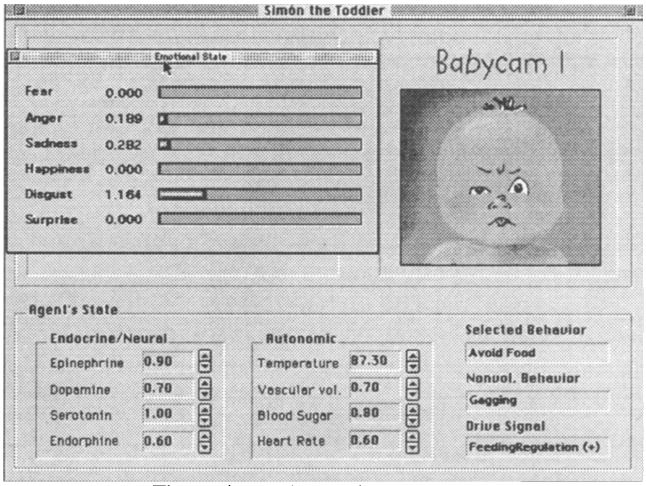

Figure 4 An Expression of Disgust

Figure 4 shows a snapshot of the environment in a situation where Simón is displaying *Disgust* right after eating *not-so-tasty* food.

Issues and Future Work

Several improvements and extensions can be made to the computational model presented in this paper. Some of the issues we are currently thinking and working on include extending the model to consider other influences of emotions in the agent's behavior (e.g. perceptual biases), and also to include memory-based elicitors of emotion, which could be used not only to generate emotions, but also to explore some of the effects of emotions in memory and learning [Bower and Cohen 1982], as well as in decision making processes, [Damasio 1994].

Conclusions

A computational model for the generation of emotions and their influence in the behavior of synthetic agents has been presented. By drawing on ideas from different fields, Cathexis offers an alternative approach to model different types of emotions, consider their dynamic nature, and differentiate them from other kinds of affective phenomena, such as moods. In contrast to other models proposed to date, it considers both cognitive and noncognitive elicitors of emotions, and it provides a flexible and extensible way of modeling the influence of emotion on the behavior of the agent, taking into account aspects of both the expressive and experiential components of emotion.

References

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

Blumberg, B. 1994. Action-Selection in Hamsterdam: Lessons from Ethology. In *Third International Conference on the Simulation of Adaptive Behavior*, 108-117. Brighton, England: MIT Press.

Bower, G.H., and Cohen, P.R. 1982. Emotional Influences in Memory and Thinking: Data and Theory. In: Clark, M.S. and Fiske, S.T. eds. *Affect and Cognition*, 291-331. Hillsdale, NJ: Erlbaum.

Damasio, A. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Gosset/Putnam.

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

Elliot, C.D. 1992. The Affective Reasoner: A Process Model of Emotions in a Multi-Agent System. Ph.D. Thesis, Institute for the Learning Sciences, Northwestern University.

Frijda, N. 1996. Passions: Emotion and Socially Consequential Behavior. In: Kavanaugh, R., Zimmerbag, B., and Fein, S. *Emotion: Interdisciplinary Perspectives*. New Jersey: Lawrence Erlbaum

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

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

Izard, C. E. 1993. Four Systems for Emotion Activation: Cognitive and Noncognitive Processes. *Psychological Review* 100(1):68-90.

Johnson-Laird, P. and Oatley, K. 1992. Basic Emotions, Rationality, and Folk Theory. In: Stein, N. L., and Oatley, K. eds. *Basic Emotions*, 169-200. Hove, UK: Lawrence Erlbaum.

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

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

Ludlow, A. 1980. The Evolution and Simulation of a Decision Maker. In: Toates, F., and Halliday, T. *Analysis of Motivational Processes*. London: Academic Press.

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

Maes P. ed. 1991. *Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back.* MIT Press.

Minsky. M. 1986. *The Society of Mind*. New York: Simon & Schuster.

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

Panksepp, J. 1994. Basic Emotions Ramify Widely in the Brain, Yielding Many Concepts That Cannot Be Distinguished Unambiguously... Yet. Ekman, P., and Davidson, R. J. eds. *The Nature of Emotion*. New York: Oxford University Press.

Pfeifer, R. 1988. Artificial Intelligence Models of Emotion. In: Hamilton, V., Bower, G. H., and Frijda, N. eds. *Cognitive Perspectives on Emotion and Motivation*, 287-320. Netherlands: Kluwer.

Picard, R. 1995. Affective Computing, Technical Report No. 321. MIT Media Laboratory.

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

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

Toda, M. 1993. Emotions and Urges. Technical Report, No. 93-1-01, School of Computer and Cognitive Sciences, Chukyo University.

Tinbergen, N. 1969. *The Study of Instinct*. London: Oxford.

Tomkins, S. 1962. *Affect, Imagery, consciousness: Vol.1 The Positive Affect*s. New York: Springer.

Velásquez, J. 1996. Cathexis: A Computational model for the Generation of Emotions and Their Influence in the Behavior of Autonomous Agents. S.M. Thesis. Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.