Emotion Synthesis: Some Research Directions

Christine L. Lisetti
Department of Information Systems
University of South Florida
Tampa, Florida 33620 USA
lisetti@coba.usf.edu

Abstract

Synthesis, needs to be presently understood as close to emotion generation: to give computers some level of understanding about what it might be *like* to have emotions. Generating states for a computer system with similar properties and/or functions to human emotions, might be one of the first steps in building such an understanding. We propose a computational design for modeling emotion generation both at the physiological and at the subjective experience level.

Introduction

Given the strong interface between affect and cognition described in an earlier paper (Lisetti et. al 1998), and given the increasing versatility of computer agents on the other hand, the attempt to enable our computer tools to acknowledge affective phenomena rather than to remain blind to them appears desirable (Lisetti et. al 1998). While research on affect recognition is contributing to this effort, the question about what to do with the data while or after it is sensed is an important one. One approach is to develop agents with emotion synthesis capabilities such that they can respond appropriately to humans. Synthesis, needs to be presently understood as close to emotion generation: to give computers some level of knowledge about what it might be *like* to have emotions. Generating states for a computer system with similar properties and/or functions to human emotions, might be one of the first steps in building such an understanding.

The present paper aspires to give a framework for the development of a multi-level computational model in which two major systems associated with emotion – physiological arousal and subjective experience – are all partially present, and their circular interactions are specified. In the remaining of this paper, we raise some issues that need to be addressed to implement emotion synthesis in an artificial agent. We propose a design for implementing such a system using a connectionist network, and we present preliminary results of a case study that we tested.

Computational Modeling

According to Pfeifer (1988), one of the limitations of the previous AI approaches to emotion modeling was that they left out the physiological component of emotional states. Ortony (1988), on the other hand, points to the lack of explicit consideration of the subjective component of the emotional experience.

AI approaches to emotion have, for the most part, been based upon Paulhan's (1887) conflict theory, in which emotions are thought to occur when an ongoing tendency is interrupted. The conflict approach emphasizes the need to simulate systems with limited resources in an unpredictable world, and with multiple goals and plans which can conflict with each others, and which, therefore, must be able to be interrupted. Simon's (1967) argument that emotions have a counter part in computational systems that work with multiple goals in finite time with limited resources is indeed related with Paulhan's theory. A number of approaches have already made a number of important contributions and have been surveyed and explained in details (Picard 1997).

Design Issues

In neuroscience, theories which study the evolutionary component of the nervous system as depicted by MacLean (1990), suggest that the key to understanding the variety of emotion processes and their relation with cognitive processes will emerge by taking an evolutionary approach to the analysis of human brain's emotional control systems. For example, Derryberry and Tucker (1992), suggest associating: (1) expression and instrumental functions with the earlier *brain stem systems* – the ANS, endocrine, and motor/expression systems; (2) increased sensitivity and flexibility to emotion signals in the environment with the *limbic system* – amygdala, hippocampus, and hypothalamus –

which evolved later; (3) more complex emotions involving cognitive processes such as salience and motivation with the *paralimbic* and *neocortical structures*, which evolved last in the brain.

In the same manner that the question arises as to whether we need to consider multiple physiological interacting systems for different emotions – rather than one system for the variety of possible emotions – computational models of emotion may need to model activity at different levels possibly using different implementations, i.e. they may need to be "hybrid" systems.

Hybrid Systems

Following Pfeifer's insight (1988), we give suggestions for the design of a hybrid model. A hybrid model mixes (1) system-theoretic considerations at the *knowledge level*, i.e. specifications of the kind of knowledge used in the system but no specifications of the processing details; (2) specification of how the knowledge is processed at the *processing level*; (3) specification of activation mechanisms which enable control of sequential and parallel processing at the *microscopic level*; (4) *implementation* of the above information into an integrated computer system.

The methodology involves the investigation of architectural principles at various levels such that emotional phenomena *emerge* from the interaction of the basic components. The design of the model is inspired from overall representation of affect described by Zajonc and Markus (1984). As shown in figure 1, emotion elicitation depends upon biological, sensory and cognitive inputs. Emotion generation depends upon gating processes, such as conflicting existing emotions, conscious and unconscious suppression, previous muscular engagement, and the like. Emotion generation is associated with three components (the boxes on the far right in figure 1) : (1) ANS arousal and visceral activity; (2) motor expression of emotion; (3) subjective experience of emotion which requires the mediation of the cognitive system.

ANS Influence

Although emotions are associated with more than one physiological systems, – the early brain stem, limbic, and neocortical systems – in this part of our system, we are presently interested in the somatic states associated with emotions. We consequently focus on modeling the ANS which was found to be the key to acheiving the appropriate modification of physiological parameters in the body (Damasio 1994).

ANS activity has long been considered undifferentiated (Cannon, 1927; Schachter and Singer, 1962; Mandler, 1975). Mandler (1975) adapted Schachter and Singer's theory (1962) in which emotions are formed

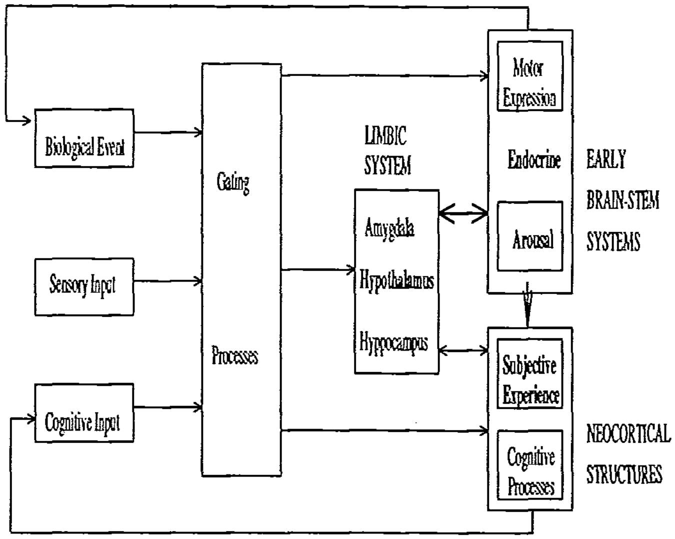

Figure 1: "Proposed Model"

from two independent systems, namely sympathetic nervous system (SNS) arousal along the physical dimension and evaluative cognition along the social dimension. It stresses the role of *interrupts* in emotion and proposes that increase in SNS activity follows the interruption of well-organized behavior. It is considered that this increase in SNS activity determines how important the emotion is, i.e. its intensity. Following emotion generation, sympathetic activity can subsequently increase depending on the level of *surprise* that the interrupt brought about and on the *intensity* of the emotion.

Other emotion theories, however, claim that different emotions are accompanied by *specific* autonomic patterns (Ekman, et.al. 1983; Izard, 1977). Recent studies on multiple receptor subtypes in both the sympathetic (Alquist, 1948) and the parasympathetic (Mitchelson, 1988) branches of the ANS, and in the various neurotransmitters and neuropeptides responsible for transmission and modulation, suggest that the ANS might have a much greater capacity for differentiated action on emotion phenomena than previously anticipated (Levenson, 1992). We model the notion of specificity, i.e. that different emotions are associated with different patterns of ANS activity.

It is important to note that most emotion theorists agree at least on the importance of the ANS in determining intensity and valence of an emotion. Zajonc's work on thermoregulation (1994) associates low body temperatures (linked with parasympathetic activity) with hedonic states, and high temperatures (linked with the sympathetic activity) with negative emotional states.

This perspective has significant implications for models of emotion elicitation, with direct relevance to artificial systems: ANS differentiation in emotion implies that changes in the environment, perceived and appraised at a simple "automatic" level, lead to the occurence of an emotional state, which then organizes and calls subordinate response systems. We based our model on this recent hypothesis that the ANS can *initiate* emotion.

This view is supported by Levenson (1992) who expects "that it will eventually be established that any component of emotion can assume this initiating role" (Levenson, 1992). As of today, however, such plasticity in elicitation has only been considered by a few, including Camras' extension (1992) of dynamical systems theory (Kugler et al., 1980; Kelso, 1995). Zajonc's suggestion of a link between facial action, hypothalamic temperature, neurotransmitter release, and subjective changes seems to support this theory further (Zajonc, 1994).

Before discussing the proposed physiological subsystem, however, we explain how the representation of subjective experience might be integrated with other subsystems.

Subjective Experience Subsystem

The subjective experience of emotion involves the cognitive system, and might be best represented by abstract structures similar to the ones described in schema theory (Schank, 1977). They derive from a transformation of sensory and kinesthetic input (Zajonc and Markus, 1984). For more complex emotions, they derive from cognitive processes of appraisal (Ortony, 1988).

Our design consists of working at the knowledge level with computational schema in a similar fashion to the computational schemas for emotion concepts derived from Wierzbicka's work (1992) and described in Lisetti (1997). These data structures are an attempt to isolate various emotional components relevant for specific emotions. They represent the "knowledge" found in a subjective emotional state and include physiological representations such as intensity, and tempo, as well as beliefs and other cognitive components. Not every emotion requires the same components. We give one example below and more examples can be found in Lisetti (1997).

```
SEMANTIC META-DEFINITION
 X feels something
   sometimes a person thinks something like this:
   something good happened
   I wanted this
 because of this,
   this person feels something good
 X feels like this
  this person feels something good
 X feels like this
```

PLEASED SCRIPT	
Causal Chain:	Emotion Components:
	User = *Tony*
	Facial Expression = *Happy*
	User = *Tony*
Feel	Facial Expression = *Happy*
↓ Think	Tempo = *fast*
↓ Happened	Intensity = *unspec.*
↓ Wanted	Involvement = *passive*
Feel	Comparison = *match*
	Chunk size = *unspec.*
	Criteria = *good*

The interface of this subsystem with the physiological subsystem described below is only partial, given that current data about ANS arousal describes the ANS as giving information about valence and intensity only (Mandler, 1975). New research on the variety of neurotransmitters associated transmission and modulation, as well as their uneven distribution along the ANS (Levenson, 1992) might lead to more specific parameters passed along between the ANS and subjective experience representations.

Physiological Subsystem

In this section, we describe the model that we have implemented, and show some of our preliminary results. With the connectionist approach, *simple* non-intelligent constituents (like neurons) express *global properties* when *connected*. Knowledge or experience is stored in the strength of the connections between units. Each constituent in such a system functions only in its local environment so that the system is not directed by a central process. However, a global cooperation emerges spontaneously (due to the configuration of the system) when the state of each 'neuron' reaches a satisfactory status.

We model how emotional patterns emerge from the neural activity along the *ANS*, from either one of its two branches, depending upon which branch is active. The *ANS* includes two branches of nerve fibers, the *parasympathetic* and *sympathetic* branches, both of which act as a seesaw, where the two branches have opposing effects on each other. The *sympathetic* branch is responsible for the arousal state of emotions such as fear, anger, and the fight-or-flight response, while the *parasympathetic* branch is responsible for the opposite appeasing responses. Both branches consist of fibers to and from the brain and spinal cord.

We represent the network as a set of artificial neurons along the *ANS* corresponding to various body

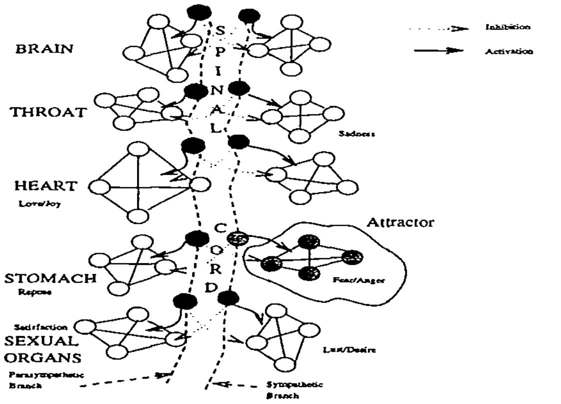

Figure 2: "Arising of an Emotional Attractor along the *ANS*"

areas such as the stomach, the heart, etc. (see figure 2). We have designed an arbitrary mapping of various emotions corresponding to the arousal of nervous activity in a particular body area. Fear, for example is associated with the stomach while sadness is associated with the throat (see figure 2). At the macro-level of description, when many interconnected units become activated simultaneously, they form a collection of active units refered to, from now on, as *attrators.*

When an attractor emerges from the activity of the network as a whole, it can be understood as modeling a particular emotional state associated with a type of activity of the ANS. As we explained earlier, ANS differentiation implies that changes of the environment, perceived and appraised at a simple automatic level lead to an emotional state which then calls subordinate response systems. To account for the responses in our model, we have included units which stand for the environment itself.

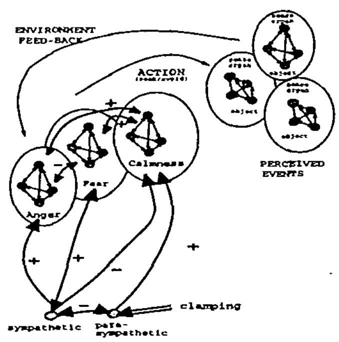

Figure 3: "Detail of units connectivity in one body area"

The detail of the network units connectivity for one particular body area, the stomach, is shown in figure 3. The stomach has been associated with the emotions of fear or anger, and calmness. The connections *from* the units of the body areas *to* the perceived environment need to be understood as *actions* taken upon the environment. The connections in the reverse direction, can be understood as the *feed-back* of the environment in response to the taken action. The connections between two possible distinct emotional attractors (such as anger and fear in this example), are mutually inhibiting. The two units for the parasympathetic and sympathetic branches are mutually exclusive and have negative inhibiting connections. The sympathetic unit is exciting to the two emotional attractors corresponding to fear and anger, while it is inhibiting to the calmness attractor. The parasympathetic unit, to the contrary, excites the calmness response with a positive connection.

The choice of mutual exclusion was made in accordance with Clyne's exclusivity principle of pure emotional states which suggests that we cannot express one emotion when we are feeling another . While testing the various emotion theory as suggested by Picard (1997), another test could be made in the future to support Plutchik's notion (1980) that emotions are rarely experienced in their pure form but are rather a mixture of some of the principal of basic emotions . Altering the connectivity matrix of the model could account for computationally testing both theories.

Some particular tendencies toward a specific pattern of activity, which models tendencies toward certain emotions, can be set at the initial state. This is accomplished by pre-setting the strength of connections (the weights) between 'neurons'. When the weights are pre-set to a high value between a collection of units standing for a certain region of the body (such as the stomach, or other), the tendency of the system is to arise and to settle in that region. This option was allocated in order to allow for predispositional emotional tendencies, some possibly being genetically encoded.

A Learning *Boltzmann* machine

In the present model, the propation rule calculates the net input of a unit by a summation of the weights multiplied by the activation values of each units connected to the unit whose activation level is being updated. It is given by the following equations:

$$net_i(t) = \sum_{j=0} w_{ij} a_j(t) \text{ where}$$

$$a_i(t) = \begin{cases} 1 & \text{if } randomnumber < \mathbf{sigmd}(net_i(t)) \\ 0 & \text{otherwise} \end{cases}$$

and where

$\text{sigmd}(net_i(t)) = \frac{1}{1+e^{-net_i(t)/T}}$

We then can get a measure of the degree to which the overall network has acheived a "good" interpretation. This measure is defined as the *goodness function*, G, where:

$$G = \sum_{ij=0} w_{ij}\, a_i\, a_j$$

In general, the model moves in such a way as to maximize *goodness*.

Furthermore, some of the connections between 'neurons' are modifiable so as to allow learning to happen. The effect of learning can be two-fold: emphasize a particular 'emotional attractor', or attract the system into a new one. The learning option was added to the system in order to account for the possibility of changing the emotional setup of the system. The relearning can happen in the human *ANS* via several processes. The *ANS* is influenced by naturally present neurotransmitters, and it can be stimulated with chemically similar compounds.

A *Hebbian* learning rule was added to the *Boltzmann* architecture to allow the network to learn without a 'teacher' such that the patterns can organize by themselves . In this manner, the emotional patterns get naturally reinforced via *Hebbian* learning (similarly to the real neurons in the body), so as to depict the build-up of emotional arousal. It is worth noting that *Hebbian* learning is atypical of *Boltzmann* machines which are usually implemented with learning with a target (or 'teacher'). This choice was made so that the model could account not only for predispositional tendencies toward one or more emotion, but also for some new patterns which may incidently occur and get reinforced. It is important to note that we later provide for unlearning and relearning so that the model can change some pre-established and/or reinforced patterns of emotional arousal later in its life-cycle .

The *Hebbian* rule used in this present model is given by the following equation:

$$\Delta w_{ij} = \eta(4a_i a_j - a_i - a_j)$$

where Δw_{ij} is the amount by which the weight connection between unit u_i and unit u_j is incremented in order to reflect the learning; and η is a parameter which can be adjusted. The effect of the rule is to strengthen the connection between two units which are on simultaneously, to decrease the connections if one unit is on and the other is off, and to keep the weights unchanged if both units are off.

Modeling Cases and Results

The system illustrates the cycle from reflex-like emotional responses to chosen ones by modeling three chronological cases described as follows: (1) some predetermined dispositions toward a particular emotion; (2) the re-inforcement of these dispositions;(3) the arising of a new interpretation aimed at altering the predispositional emotional patterns.

Case One: Pre-dispositional Tendencies

This portion of the modeling (see figure 4) provides an illustration of how the system can be perturbed from *within*. Inner perturbations constrained by the nervous system's organization (rather than external inputs) are presently considered to be the cause of emotional arousal in the human body. The choice of the *Boltzmann* machine was made accordingly so as to model the spontaneous inner activity of the nervous system.

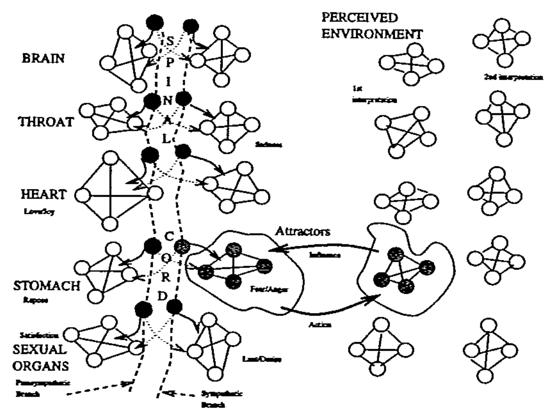

Figure 4: "Coupling with the Environment"

The inner pertubations of the network can be understood as bodily sensations, leading to the arising of a particular emotional state. This emotional state, in turn, influences the type of action that the system will take upon its environment. The loop continues further, as the actions taken feed-back onto the system itself, leading to new bodily sensations, and so on. The connections from the body units to the perceived environment units can be understood as the actions taken by the system. The connections in the reverse direction can be understood as the influence of the environment upon the system.

As explained earlier, at the initial state, some particular tendencies toward a specific pattern of activity which models tendencies toward certain emotions can be pre-set. Just as individual A might have a particular tendency toward anger, and individual B might be more prone to sadness, so our network can instantiate different agents with (so to speak) different personalities. The *degree* to which these tendencies exist can also be set depending on how strongly predisposed toward one emotional state we want our instantiation of the model to be.

Case Two: Learning

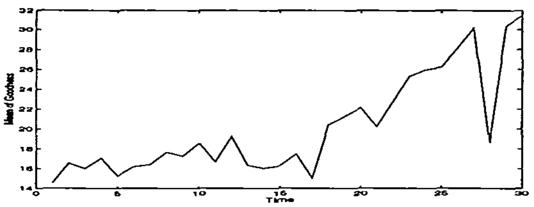

Figure 5: "Case two: plot of the mean of the *goodness function* of the learning process"

This very same model then grows into a learning system. While some e motions might not have been reflex-like or pre-wired to start with (as in the previous case modeled), these emotions can still become reflex-like for the system. Once an emotional attractor emerges, it becomes more and more attractive through *Hebbian* learning. As the system grows 'older', patterns are reinforced by the learning process and emotional reflexes built up.

The increase of the *goodness function* over time, shown is figure 5, is an indicator of how the network learns a particular pattern of activity. With a low tendency toward a particular emotional state to start with (sadness, in this particular case), the learning system still ends up in a state of equal intensity to the starting state of the previous network.

The curve of the learning process (figure 5) has a linear overall structure, although, at the 27th life-cycle, the system jumps out of its emotional state to experience a temporary lapse of some other emotion (fear in this case), and continues learning sadness.

Case Three : Unlearning and Relearning

Lastly, the system evolves into the stage characterized by the understanding of a new interpretation of the emotional arousal (see figure 6). It is here assumed that by increasing the activity of the parasympathetic branch, the system can reach a different state, as the bodily sensations are brought back to a calmer response. This means that the activity of the parasympathetic nervous system is activated and/or emphasized at a particular location of the body (in this example, the stomach area).

In the neural network, this is accomplished by keeping the parasympathetic unit on or *clamped* for a long period of time at a particular body region (see figure 3). Hence, the network evolves from a predispositional tendency toward anger (as in case one above, figure 4), to calmer bodily arousal, which leads the system to a reinterpretation of the emotional arousal of anger (see figure 6). In the future, when the anger attractor arises, the calmness attractor will arise as well, keeping the system in a state of balance where the original primal emotion points to another interpretation.

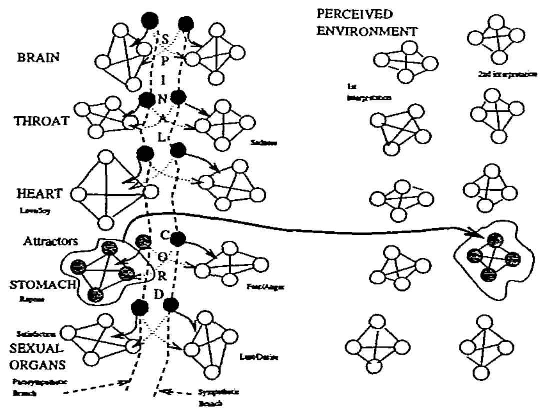

Figure 6: "Actively Coupling with the Environment"

Discussion

Further improvements could be made in the following directions: (1) The stimulation of the each branch of the *ANS* could be implemented by replacing the clamping of the *ANS* units by the activation of neuromodulators (Rumelhart 1995) ; (2) The different pathways of the parasympathetic and sympathetic branches could be further refined by connecting these units together and simulate ascending and descending pathways; (3) Various experiments with existing competing theories of emotion could be instantiated and tested by looking at the physiological and behavioral level.

Conclusion

In conclusion, the model we have developed is only a first pass toward a model of emotional states. Further work is of course required to develop a better understanding of our emotional states, and to simulate their function on artificial systems.

References

R. P. Alquist. A Study of Adrenotropic Receptors. *American Journal of Physiology*, 153:586–600, 1948.

L. Camras. A Dynamical Systems Perspective on Expressive Development. In K. Strongman, editor, *International Reviews of Studies on Emotion*. New York: Wiley, 1992.

W. Cannon. The James-Lange Theory of Emotions: A Critical Examination and an Alternative Theory. *American Journal of Psychology*, 39:106–124, 1927.

D. M. Clynes. *Sentics: The Touch of Emotions.* Anchor Press Doubleday, 1977.

A. Damasio. *Descartes' Error: Emotion, Reason, and the Human Brain.* New York: Grosset/Putman Book, 1994.

D. Derryberry and D. Tucker. Neural Mechanisms of Emotion. *Journal of Consulting and Clinical Psychology*, 60(3):329–337, 1992.

P. Ekman, R. W. Levenson, and W. Friesen. Autonomic System Activity Distinguishes among Emotions. *Science*, 221:1208–1210, 1983.

D. Hebb. *The Organization of Behavior.* John Wiley, 1949.

G. E. Hinton and T. J. Sejnowski. Learning and Relearning in Boltzmann Machines. In *Parallel Distributed Processing: Explorations in the Microstructures of Cognition Volume 1: Foundations.* Cambridge: MIT Press A Bradford Book, 1986.

C. E. Izard. *Human Emotions.* New York: Plenum Press, 1977.

S. Kelso. *Dynamic Patterns: The Self-Organization of Brain and Behavior.* Cambridge, Mass: MIT Press, 1995.

P. Kugler, S. J. Kelso, and M. Turvey. On the Concept of Coordinative Structures as Dissipative Structures: I Theoretical Line. In G. Stelmach and J. Requin, editors, *Tutorials in Motor Behavior.* Amsterdam: North Holland, 1980.

R. Levenson. Autonomic Nervous System Differences Among Emotions. *Psychological Science*, 3(1):23–27, 1992.

C. L. Lisetti. Motives for Intelligent Agents: Computational Scripts for Emotion Concepts. In G. Grahne, editor, *Proceedings of the Sixth Scandinavian Conference on Artificial Intelligence (SCAI-97)*, Helsinki, Finland, August 1997. Amsterdam: IOS Press Frontiers in Artificial Intelligence and Applications.

C. L. Lisetti, D. E. Rumelhart, and M. Holler. An Environment to Acknowledge the Interface between Affect and Cognition. In *Working Notes of the 1998 AAAI Spring Symposium Series on Intelligent Environments (AAAI SSS-98). Technical Report SS-98-02.* Menlo Park, CA: AAAI Press, 1998.

P. D. MacLean. *The Triune Brain in Evolution: Role of Paleocerebral Functions.* New York: Plenum Press, 1990.

G. Mandler. *Mind and Emotion.* New York: Wiley, 1975.

F. Mitchelson. Muscarinic Receptor Differentiation. *Pharmacology and Therapeutics*, 37:357–423, 1988.

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

F. Paulhan. *The Laws of Feeling* English Trans. 1930. London: Kegan-Paul, 1887.

R. Pfeifer. Artificial Intelligence Models of Emotion. In V. Hamilton, G. Bower, and N. Frijda, editors, *Cognitive Perspectives on Emotion and Motivation*, pages 287–320. Dordrecht, Netherlands: Kluwer Academic Publishers, 1988.

R. W. Picard. *Affective Computing.* Cambridge, MA: M.I.T. Press, 1997.

R. Plutchik. *Emotion Theory, Research and Experience, volume 1 Theories of Emotion.* Academic Press, 1980.

D. Rumelhart. Neuro-Modulation: A Connectionist Approach. In H. Morowitz and J. Singer, editors, *The Mind, the Brain and Complex Associative Systems*, pages 145-153. Reading, Mass: Addison-Wesley, 1995.

S. Schachter and J. E. Singer. Cognitive, Social and Psychological Determinants of Emotional State. *Psychological Review*, 69:379–399, 1962.

R. Schank and R. Abelson. *Scripts, Plans, Goals and Understanding.* Erlbaum, 1977.

H Simon. Motivational and Emotional Controls of Cognition. *Psychological Review*, 1:29–39, 1967.

A. Wierzbicka. Defining Emotion Concepts. *Cognitive Science*, 16:539–581, 1992.

R. Zajonc. On the Primacy of Affect. *American Psychologist*, 39:117–124, 1984.

R. Zajonc. Emotional Expression and Temperature Modulation. In S. Van Goozen, N. Van de Poll, and J. Sergeant, editors, *Emotions: Essays on Emotion Theory*, pages 3–27. Hillsdale, NJ: Lawrence Erlbaum Associates, 1994.