7 Emotion Synthesis

If computers are ever to "have" emotions, then one of the things they need is the ability to synthesize or generate them. In Chapter 2, I described five components of a system that can be said to have emotions. These were:

1. Emotional behavior;
2. Fast primary emotions;
3. Cognitively generated emotions;
4. Emotional experience: cognitive awareness, physiological awareness, and subjective feelings;
5. Body-mind interactions.

Depending on the task at hand, certain subsets of these five components will suffice. Just as all animals do not need emotion systems as sophisticated as a human emotion system, neither do all computers. Furthermore, differences in computers and humans, especially their different physiologies, imply a variety of possible interpretations for these components, especially for the fifth one.

This chapter addresses how to begin giving these abilities to computers. Earlier chapters illustrated the benefits of such abilities which, in humans, include more flexible and rational decision-making, the ability to determine salience and valence, improved reasoning ability, and a variety of other beneficial interactions with creativity, learning, attention, memory, and regulatory processes. We can expect computer emotions to play a role in giving computers these more human-like abilities, together with improving their skills for interacting with people.

One of the areas in which computer emotions are of primary interest is software agents, computer programs that are personalized—they know the user's interests, habits and preferences—and that take an active role in

assisting the user with work and information overload.[1] They may also be personified, and play a role in leisure activities. One agent may act like an office assistant to help you process mail; another may take the form of an animated creature to play with a child. The notion of an agent raises several expectations from the human user. In particular, how can agents be made to be personalized, intelligent, believable, and engaging?

"Give them emotions" is not the entire solution to these problems, but it is a critical component. The assistant that cannot read your emotional expression, reason about what your emotions might be, and learn what is important to you—when not to interrupt, for example—will act unintelligently. If the agent cannot have a mechanism for the equivalent of "feeling bad" for causing you distress, then it is likely to repeat this behavior. The lack of such a mechanism is believed to be at the root of the problem of the emotion-impaired patients who know what to do, but do not do it. An ability to "feel good or bad" does not merely effect the agent's ability to learn, but helps it prioritize and choose among all its actions—learning, planning, decision-making, and more. Chapter 2's scenario of a smart personal assistant illustrated a case where emotions in an agent were important for its ability to address multiple concerns in an intelligent and efficient way.

Emotions have been implemented in agents today, but not in this way. The emotions implemented today are primarily cognitively generated, the third component only. Furthermore, they have mostly been used only for entertainment purposes. The agents have some simple cognitive emotions, and they can usually express these emotions, but they do not have the ability to recognize the emotions of people, to experience or show empathy, or to benefit internally from the functions that emotions can provide. Instead of using emotion to help manage information overload, regulate prioritization of activities, and make decisions more flexibly, creatively, and intelligently, today's agents use emotion only to entertain. This is a fine use, and valuable for many applications, but it should not be the only use.

As we begin to construct systems that can synthesize emotions, we need to consider emotional intelligence, teaching computers how to control their emotions, when and how to express them, and how to correctly and wisely recognize and reason about emotion. These abilities are of great importance. If a system cannot handle emotions intelligently, then perhaps it should not synthesize them at all. However, emotional intelligence is hard to develop without first having a system that has emotions. I suggest that once the emotion synthesis mechanisms that I describe in this chapter are fully in place, emotional intelligence will need to be learned, probably from social interactions.

Let us begin now to consider means of giving computers the five components above. Of these five, the easiest to start giving a computer is the third, cognitively-generated emotions. I will describe this in the next section. After that, I will describe models that rely upon a combination of mechanisms for generating emotions. Finally, I will describe ways in which computers' emotions can interact with other processes, and begin to provide some of the beneficial influences that emotions exert in human decision-making, learning, behavior, and more. Along the way, I will illustrate each of these pieces with examples, including examples from the literature where they exist. The reader is referred to the overviews of Pfeifer (1988) and Hudlicka and Fellous (1996) for descriptions of additional efforts to implement various aspects of emotion synthesis in computers. I will also describe several pieces that have yet to be implemented by researchers, but which are nevertheless important for synthesizing emotion and its influences. Taken together, these pieces begin to fill in the framework needed to construct affective computers with the ability to synthesize emotions.

Emotion Synthesis via Cognitive Mechanisms

There are dozens of theories about how emotions are generated, some of which were mentioned earlier. Any emotion theory can be simulated on a computer. Indeed, the process of designing simulations is a valuable aid in developing theories, stimulating new thinking and questions. I will highlight two theories in this section that have been designed with computation in mind, and that have been given at least a trial implementation in computers. Each implements the third component of emotions—cognitively-generated—and thereby provides a key piece in the framework of affective computing.

The Ortony Clore Collins (OCC) Cognitive Model

The first theory that I will describe for emotion synthesis was never intended to be used for emotion synthesis. Nonetheless it is useful for synthesizing cognitive emotions. In 1988 Ortony, Clore and Collins published their book, *The Cognitive Structure of Emotions*, setting forth a model of cognitive appraisal for emotions that has come to be called the "OCC" model. Ortony et al. wrote that they did not think it was important for machines to have emotions; however, they believed AI systems must be able to *reason about* emotions—especially for natural language understanding, cooperative problem solving, and planning. Some structure was needed so computers could begin to represent the thicket of concepts considered to be emotions.

The OCC model addresses the problem of representing emotions not by using sets of basic emotions, or by using an explicitly dimensioned space, but, by grouping emotions according to cognitive eliciting conditions. In particular, it assumes that emotions arise from valenced (positive or negative) reactions to situations consisting of events, agents, and objects. With this structure, Ortony, Clore, and Collins outlined specifications for 22 emotion types, as given in the boxes along the bottom of Fig. 7.1. Additionally, they included a rule-based system for the generation of these emotion types.

Although the OCC model has not been fully implemented in any AI systems, it was the first model to cater to the AI community in terms of framing rules that are relatively easy to implement in computers. Despite the original intentions of Ortony et al. it has also become the default model for *synthesizing* emotions in computers, even though it only addresses cognitive emotion generation. Let us consider an example, how the emotion joy is synthesized in the OCC model:

Synthesis of Joy. Let $D(p, e, t)$ be the desirability of event e that person p assigns at time t. This function returns a positive value if the event is expected to have beneficial consequences, and returns a negative value if the event is expected to have harmful consequences. Let $I_g(p, e, t)$ represent a combination of global intensity variables (e.g., expectedness, reality, proximity.) Let $P_j(p, e, t)$ be the potential for generating a state of joy. Then an example rule for joy is:

$$\begin{aligned} &\text{IF } D(p, e, t) > 0 \\ &\text{THEN set } P_j(p, e, t) = f_j(D(p, e, t), I_g(p, e, t)) \end{aligned} \tag{7.1}$$

where $f_j(\,)$ is a function specific to joy.

Similar rules can be used for computing potentials for other emotions. For example, the potential for distress, $P_d(\,)$, is computed by changing the "IF" to test for negative desirability, and the "THEN" to use a suitable f_d instead of f_j.

The rule above does not cause a state of joy or an experience of a joy feeling, but is used to trigger another rule that sets up an intensity of joy, I_j. Given a threshold value, T_j, then:

$$\begin{aligned} &\text{IF } P_j(p, e, t) > T_j(p, t) \\ &\text{THEN set } I_j(p, e, t) = P_j(p, e, t) - T_j(p, t) \\ &\text{ELSE set } I_j(p, e, t) = 0 \end{aligned} \tag{7.2}$$

This rule activates the joy emotion—giving it a nonzero intensity—when the joy threshold is exceeded. The resulting intensity can be mapped to one of a variety of emotion terms in the joy family, such as "pleased" for a moderate value or "euphoric" for an unusually high value.

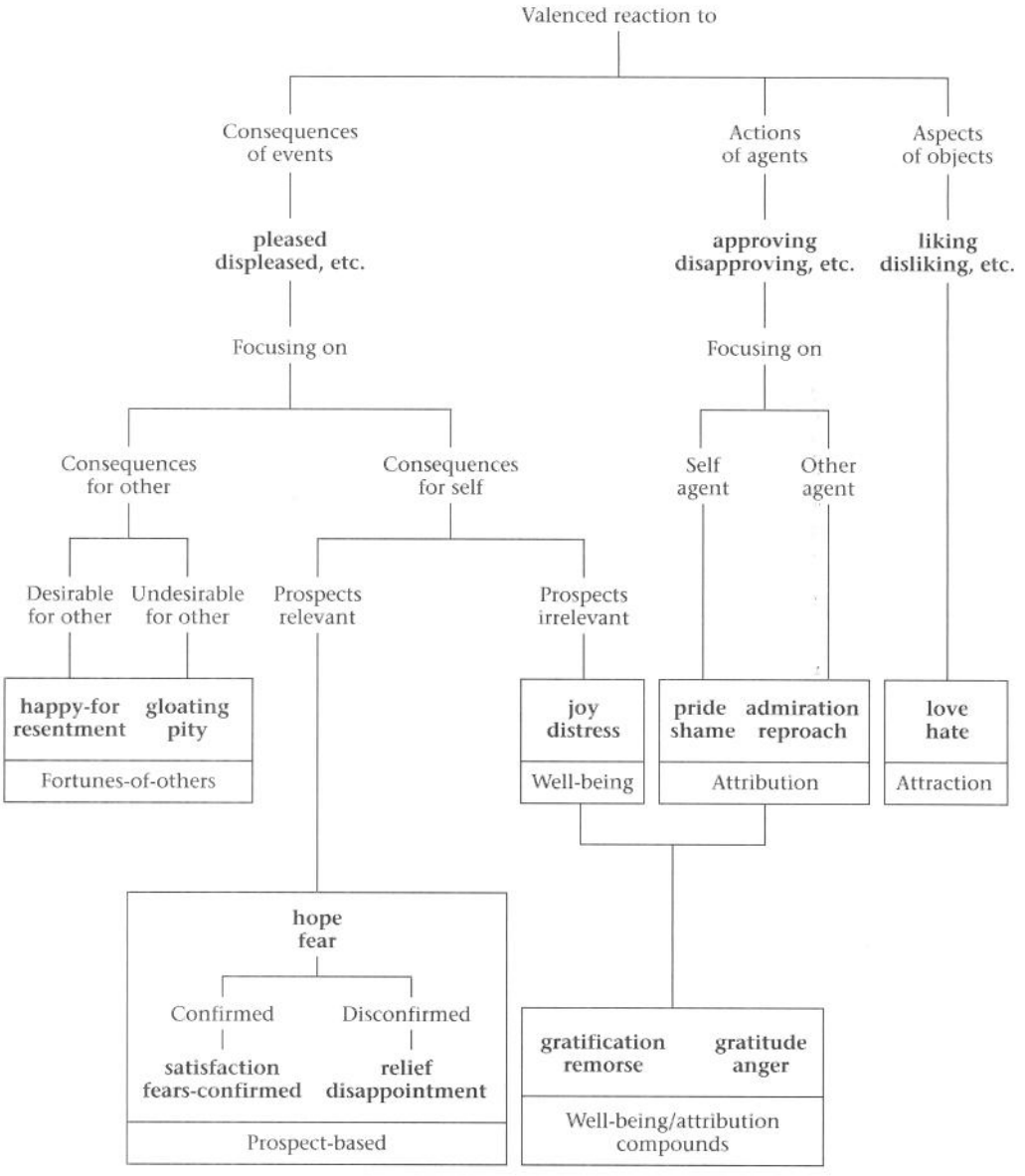

Figure 7.1
The OCC cognitive structure of emotions. (Reprinted from Fig. 2.1 of Ortony, Clore, and Collins (1988) with permission from Cambridge University Press.)

The examples of joy and distress are the simplest cases; more complicated rules exist for other emotional types in the OCC model. Ortony, Clore, and Collins omit low-level details of implementation in their model—especially with respect to how emotions interact, mix, and change their intensity with time, what values to use for the thresholds, and what form to use for functions such as f_j. However, this low level of representation can be addressed in the manner I described in Chapter 5.

The OCC model synthesizes emotions as outcomes of situations, which include events, objects and agents. Since being in an emotional state is itself a situation, the model also permits emotions to trigger additional emotions, or to repeatedly trigger the same emotion. For example, the inability to cope with a particularly intense emotional state can trigger new emotions: the long-hoped-for return of hostages causes loved ones such relief that they shed tears of joy. Thus, an overwhelming positive state can trigger an emotional expression usually associated with a negative state. Another example arises when an inability to cope with a negative state causes additional negative emotions: Rhonda is trying to learn to control her anger, and finds herself as angry as ever at something somebody did that is beyond her control. Upon reflection, she becomes angry at herself for letting herself become so angry, thereby intensifying her anger all the more. Negative emotional situations can also trigger positive emotional states. Ortony et al. did not write about this case, but here is an example: Chris has difficulty expressing emotions, and was taught in his childhood that it is weak to cry. Years later, upon the death of a loved one, he learns that it is healthy to cry while grieving. However, he has trouble letting himself cry. When he finally lets the tears flow, he not only feels better because of the release of some of his grief, but he feels better that he overcame his inability to cry. His tears feel doubly good. The OCC model as implemented by Elliott, illustrated below, handles cases like this.

The OCC model is not just useful for reasoning about emotions and for cognitive generation of emotions, but it also can be used to trigger other important emotional consequences—such as the subjective experience of feeling an emotion, or an emotional valence, positive or negative, to attach to a situation, so that it is more likely to be recalled when the person is in a mood congruent with that valence. As described earlier, these are important aspects of emotions in humans; I will say more about implementing them later.

Poker-Playing Agents with Facial Expressions

Earlier I described the success of using facial expressions on poker-playing software agents, and Koda's results which indicated that people preferred to play with the agent that was facially expressive. The emphasis in these

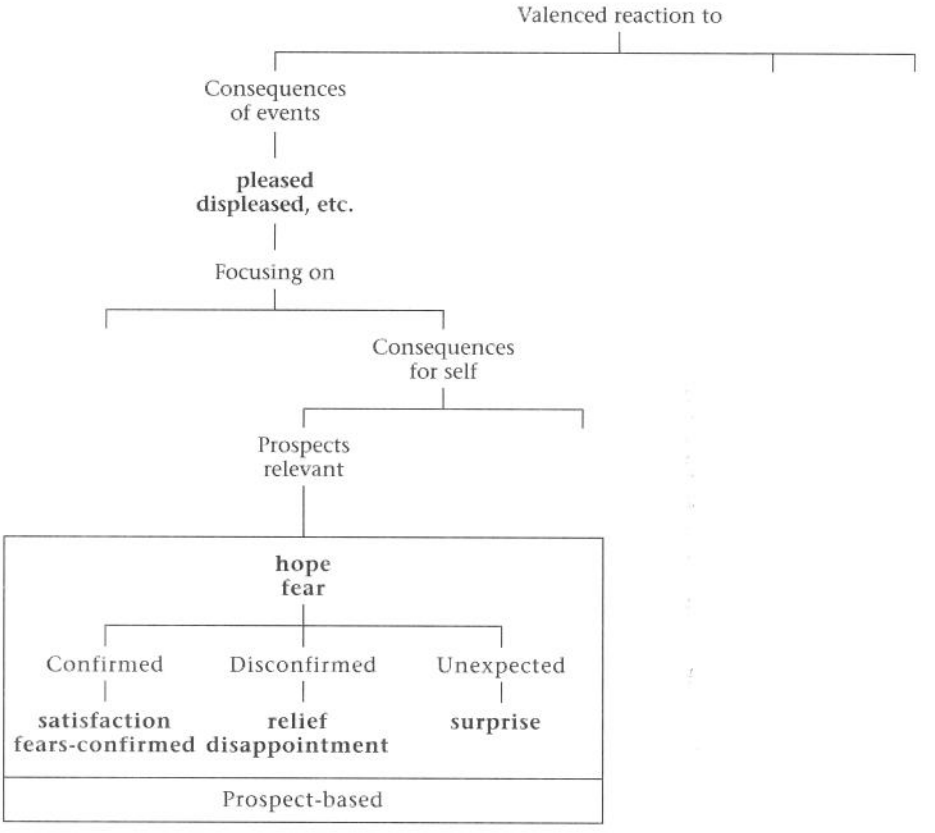

Figure 7.2
Structure used to synthesize emotional states in poker-playing agents. (Figure from Koda (1996), used by permission.)

experiments was not on emotion synthesis, but rather on generating facial expressions in situations where an underlying emotion model determined what would be expressed. However, Koda's work provides a relatively simple situation for illustrating how emotions can be synthesized using the OCC model.

Ten emotional expressions were permitted for each agent: neutral, pleased, displeased, excited (hope), very excited (hope), anxious (fear), satisfied, disappointed, surprised, and relieved. The underlying emotional states were determined with a modified subset of the OCC model, as shown in Fig. 7.2. Although the poker scenario could use the full model, Koda limited this experiment to emotions provoked only by events that have self-consequences. This particular branch of the OCC model was then augmented by adding a surprise state (such as when a player wins unexpectedly) since the OCC model does not include surprise.[2]

The poker situations giving rise to each emotion are shown in Fig. 7.3. Consider an example event of drawing a very good hand. The self-consequence of the event would be the emotion "pleased." Next, according to the OCC model, the person considers prospects for himself. In the poker game, this occurs during the betting phase, where the poker player may feel "excited" about the prospects of winning. When the game is over, if the player has won, then his hopes are confirmed and he may feel "satisfied." Of course, other outcomes are possible, and are determined by the rules of the OCC model, with the minor modification to allow for surprise which might occur, for example, if the player has a bad hand, decides not to bluff, bets anxiously, and winds up winning.

Emotions and Moods for Animated Characters

Researchers Joe Bates and Scott Neal Reilly, of Carnegie Mellon, have been interested in making agents *believable*, giving them the illusion of life. This is the goal of their "Oz project," which contains a variety of synthetic characters that may not look like any real creatures, but that are designed to be able to powerfully influence their audience as if they were real (Bates, 1994). Bates and other researchers interested in believable agents have turned to the most successful animators of all time for their answer to what provides the "illusion of life." In the magical *Disney Animation* (Thomas and Johnson, 1981), the Disney masters emphasize the importance of each character having a clear emotional state at all times. They describe numerous techniques for accomplishing this, arguing that the portrayal of emotions is what gives the Disney characters the illusion of life.

Inspired by the Disney animators, Bates and his colleagues have created emotions for their animated creatures, together with a host of tools that assist artists in building emotions for characters. One of their creatures is a house cat named Lyotard, which has a large repertoire of emotions and corresponding behaviors. For example, Lyotard can *hope* to be fed, and can be *pleased* when food is provided, and might *purr* or *rub against someone* when it is happy. The underlying emotion generation system for the Oz characters is "Em," which is part of a broad architecture called "Tok." The full architecture integrates not just emotions, but also rudimentary perception, goal-directed behavior, and language. A description of Lyotard and the Tok architecture is provided by Bates (1992). The focus of our interest is on Em, which generates the emotions for Lyotard and other characters.

Em is equipped with a default emotion system that is based on the OCC model, and hence emphasizes cognitive appraisal for emotion generation. Em is also augmented with mechanisms for generating some primary emotions, such as startle, although the way it is implemented is not distinguished

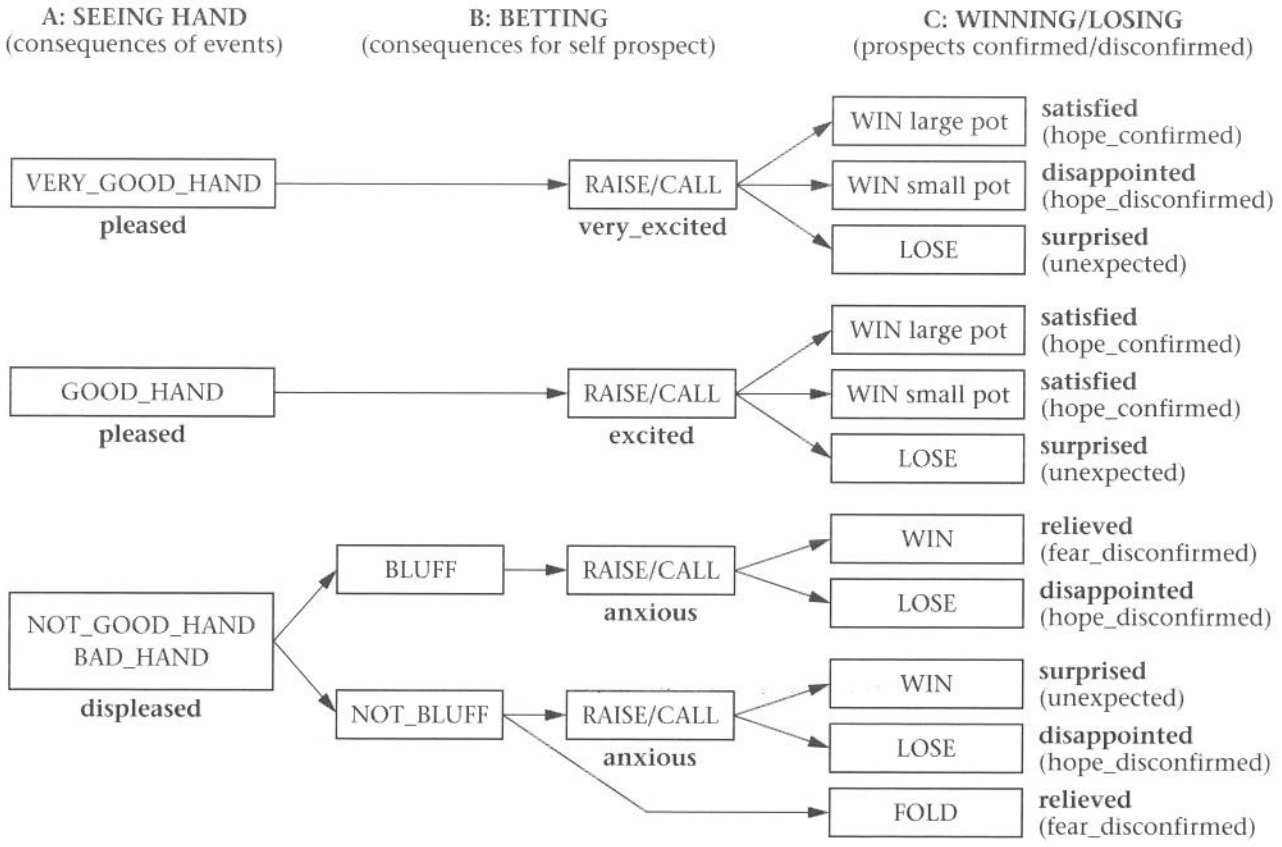

Figure 7.3
Agents emotions are generated according to these poker events. (Figure from Koda (1996), used by permission.)

from the way the other emotions are implemented. In Em's default emotion system, the emotions have intensities that are influenced by the importance of the goal that generated them. Each emotion also has a threshold, and only when its intensity exceeds this threshold does the emotion influence any outward behavior. Em also explicitly models emotion decay, where each emotion has its intensity lowered every clock cycle, until the intensity is zero. The artist is free to modify the thresholds and choose the method of decay. In these ways, Em implements several of the properties described in Chapter 5.

Most importantly, Em's emotions are arranged in a hierarchy, shown in Fig. 7.4, which separates the positive and negative, making it easy to determine states such as *good-mood* and *bad-mood*. Mood is determined differently here than the way I proposed in Chapter 5: First, Em combines all the top-level positive emotions, e.g., joy, hope, happy-for, etc., summing their intensities as follows:

$$I_p = \log_2\left(\sum_e 2^{I_e}\right), \qquad e \in \{\text{positive emotions}\}$$

Em repeats this for the set of negative emotions, to form I_n, a combined intensity for the negative emotions. If $I_p > I_n$ then *good-mood* is set to I_p and *bad-mood* is set to zero. Otherwise *good-mood* is set to zero and *bad-mood* is set to $-I_n$. The "either good or bad" mood this provides is a wise default artistically, since it is usually considered important for a character to clearly communicate one thing at a time.

Emotions generated by Em influence some cognitive activity such as the generation of new goals as well as behavior (Neal Reilly, 1996). For example, in an office situation, one character might become so angry at another that she generates a goal to get revenge. Another character might be so happy that he generates a goal to go dancing. Emotions can also influence perception. This is implemented, for example, in the graphical "Woggles" characters, where one woggle that is angry and sees two others bouncing around will likely perceive their behavior as fighting, whereas a different perception would occur if the observing woggle was not angry.

At present, much of Em's rules and cognitive and behavioral influences are hard-coded and are changed by hand by an artist or programmer to adapt to new characters and situations. These include social rules of interaction, which one might argue should be learned, not hard-coded, in a natural model of emotions. However, in the Oz project, a goal is to give the artist deliberate control over the character and its development. If rules and emotional influences were to be learned, then the artist would lose some of this control. Nonetheless, the fact that the Oz characters have social interactions that both influence and are influenced by emotions is an important step.

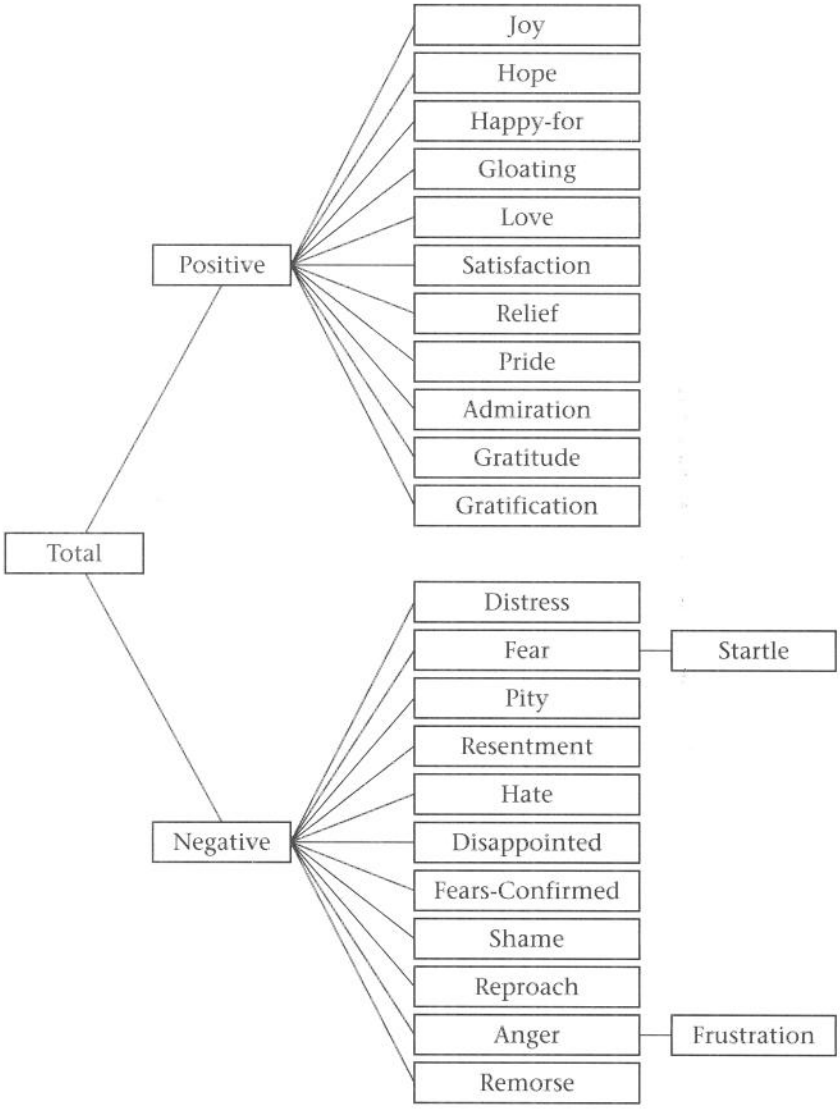

Figure 7.4
The default hierarchy of emotions in Em. (Reprinted from Figure 4-1 of Neal Reilly, 1996, by permission.)

One of the points to remember in emotion synthesis is that emotions do not completely determine actions—they only influence them. Other factors such as the type of creature and environment (e.g., aggressive wolf in the wild vs. nerd on the playground), its personality characteristics, its values, and so forth, work with emotions to influence behavior. The Tok and Em architectures provide tools for artists to manipulate these many influences.

Emotions in Social Relationships

Let's look at one final illustration of how the OCC model can be modified and used in emotion synthesis, with a different emphasis—generation of emotions among characters with social relationships. Clark Elliott of DePaul University has augmented the OCC model from twenty-two to twenty-six emotion types, and used these as the basis of a system for synthesizing and recognizing emotions based on cognitive reasoning. Table 7.1 summarizes the conditions required to synthesize each of the twenty-six emotion types. These conditions are implemented as rules in Elliott's "Affective Reasoner" system. Based on these rules, a software agent encounters conditions which can elicit the twenty-six emotion types.

The Affective Reasoner demonstrates how modeling personalities of agents and their social relationships can interact with the generation of emotions. Elliott models personality in two parts. The first part addresses how events, acts, and objects are interpreted with respect to an individual agent's goals, standards, and preferences. For example, when the winning shot of the game is scored, two agents might feel differently about the arrival of the end of the game. One might feel sad to have lost the game; another might feel happy to have finally gotten to play. The second part addresses how an agent will act or feel in response to an emotional state. An agent with an outgoing personality might express her joy verbally. A more quiet type might simply enjoy an internal feeling of happiness. This part of personality might be thought of as influenced by temperament.

Agents models three kinds of social relationships and their influences on emotions:

- Friendship. An agent will tend to have similarly valenced emotions in response to the emotions of another agent.
- Animosity. An agent will tend to have oppositely valenced emotions in response to the emotions of another agent.
- Empathy. An agent will temporarily substitute the presumed goals, standards, and preferences of another agent for its own. It will then synthesize emotions based on these presumed goals, standards and preferences, in an effort to feel what it thinks the other agent would feel.

In order for an agent to have empathy, and other emotions based on users and other agents, it maintains an internal representation of the presumed ways in which others appraise the world. This internal representation allows it not only to generate empathic responses, but also to generate fortunes-of-others emotions such as gloating. As I described earlier, responses such as empathy are key components of emotional intelligence.

Table 7.1
Emotion types used in the Affective Reasoner, based on the OCC model. (Table courtesy of Clark Elliott.)

Group	Specification	**Name** and Emotion Type
Well-being	appraisal of a situation as an *event*	**joy**: pleased about an *event* **distress**: displeased about an *event*
Fortunes-of-others	presumed value of situation as an *event* affecting another	**happy-for**: pleased about an *event* desirable for another **gloating**: pleased about an *event* undesirable for another **resentment**: displeased about an *event* desirable for another **jealousy**: resentment over a desired mutually exclusive goal. **envy**: resentment over a desired non-exclusive goal. **sorry-for**: displeased about an *event* undesirable for another
Prospect-based	appraisal of a situation as a prospective *event*	**hope**: pleased about a prospective desirable *event* **fear**: displeased about a prospective undesirable *event*
Confirmation	appraisal of a situation as confirming or disconfirming an expectation	**satisfaction**: pleased about a confirmed desirable *event* **relief**: pleased about a disconfirmed undesirable *event* **fears-confirmed**: displeased about a confirmed undesirable *event* **disappointment**: displeased about a disconfirmed desirable *event*
Attribution	appraisal of a situation as an accountable *act* of some agent	**pride**: approving of one's own *act* **admiration**: approving of another's *act* **shame**: disapproving of one's own *act* **reproach**: disapproving of another's *act*
Attraction	appraisal of a situation as containing an attractive or unattractive *object*	**liking**: finding an *object* appealing **disliking**: finding an *object* unappealing
Well-being / attribution	compound emotions	**gratitude**: admiration + joy **anger**: reproach + distress **gratification**: pride + joy **remorse**: shame + distress
Attraction / attribution	compound emotion extensions	**love**: admiration + liking **hate**: reproach + disliking

To illustrate the influence of social relationships on emotions, consider the following dialog between two agents (Elliott, 1994):

```
(The agent has previously learned that the word "Ruth" is a
female friend of the user)

User: "Ruth is unhappy"
Agent: "I am sorry to hear that. You must feel sorry for her."

(days later)

Agent: "How is Ruth?"
User: "Ruth is jealous of her rival."
Agent: "How much is Ruth feeling this resentment?"
User: "Very much."
Agent: "Perhaps you are worried about Ruth?"
```

Elliott acknowledges that the discourse here is not terribly engaging; instead, the emphasis is on reasoning about emotions within a social context. In this scenario, the agent reasons that since the user and Ruth are friends, and Ruth is sad, then it is plausible that the user feels *sorry-for* her. This is an instance of the emotion type "pity" in the OCC model. Similarly, the Affective Reasoner knows that jealousy maps to the negative emotion type "resentment" and that when Ruth is in a highly-negative state, a friend might be *worried* about her. In the OCC structure, the emotion type "fear" contains the emotion "worried" as a low-intensity instance.

Two other significant aspects of the Affective Reasoner are (1) Its forward logic-based reasoning from presumed appraisals, and events, to guesses about the emotions of others, and (2) Its backward, case-based, reasoning from facts about the situation and expressions of other agents, to the presumed emotions of other agents, and hence to the presumed appraisals of other agents. An agent might ask, "What cases do I have on file for THIS agent? for agents LIKE this agent? for agents in general?" and lastly, "how would *I* feel if these tokens were present?" These aspects are important for giving computers the ability to recognize emotions, not based on patterns of expressions as I described in the last chapter, but based on higher-level reasoning about how circumstances tend to give rise to emotion.

Roseman's Cognitive Appraisal Model

One of the newest appraisal theories, which shows promise for computer implementation of cognitive emotions, is that of Ira Roseman, at Rutgers University. Roseman has developed a categorization of the appraisals people make about events that cause emotions. Roseman and his colleagues have

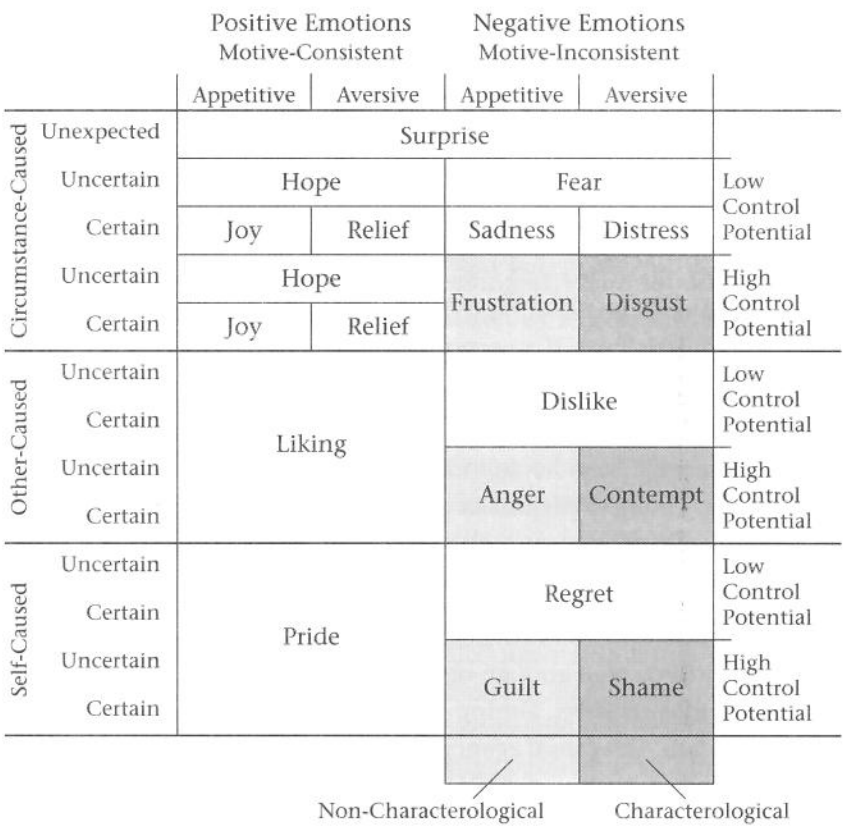

Figure 7.5
Roseman's structure for cognitively elicited emotions. (Reprinted from Fig. 2 of Roseman, Antoniou, and Jose (1996) with permission.)

run a series of studies in which subjects either recalled emotional experiences and answered questions designed to measure the appraisals leading up to the emotions, or in which subjects read brief stories of situations happening to protagonists, and answered questions about what emotion they thought the protagonist would feel, and its intensity. From these studies, Roseman and his colleagues constructed a model in which a small number of appraisals interact to give rise to seventeen emotions (Roseman, Antoniou, and Jose 1996). The six appraisals are summarized in Fig. 7.5; they are:

1. Unexpectedness. This singularly elicits surprise.

2. Motivational State and Situational State. Does the individual aim to get a reward (appetitive motive) or to avoid a punishment (aversive motive), and does the situation fit the person's motive? Situations consistent with an appetitive motive (getting a reward) elicit joy; situations consistent with an

aversive motive (not getting punishment) produce relief. Situations inconsistent with an appetitive motive (not getting a reward) elicit sadness; those inconsistent with an aversive motive (getting punishment) produce distress.

3. Probability. Is the outcome certain or uncertain? Hope and fear, unlike joy, relief, distress, and sadness, tend to follow from uncertainty.

4. Control Potential. When a negative event occurs, does the individual believe that he or she has the potential to control it? If so, frustration or disgust result, depending on the next appraisal, problem type.

5. Problem Type. If an event is negative because it blocks a goal, frustration is experienced. But if something is perceived as negative intrinsically (in its essential character), then disgust results.

6. Agency. Emotions felt toward people are produced if an event is seen as caused by other persons or the self, and one thinks about the agent. Events attributed to someone else elicit liking-love, dislike, anger, or contempt, whereas events attributed to the self elicit pride, regret, guilt, or shame.

Consider the following example: John aims to earn an A, but it is uncertain if he will do well enough on the final exam to receive one. His motivational state is appetitive, aiming for a reward. His situational state is presently uncertain. The causal agency is a test (impersonal). He has been working hard and thinks he has potential to receive an A. His appraisal of his situation suggests that he feels *hope*. If he then receives his grade and it is not an A (motive-inconsistent), then he may feel *frustration*. If he feels that his failure on the test was due to the professor grading him unfairly, then he is likely to feel *anger* toward that professor.

This model suggests that appraisals are influenced by shifts in attention. If Jill wants an A and gets one, she may focus on the A and feel joy. Or, she may think about the teacher and feel liking for him, or she may focus on what she accomplished, herself, and feel pride. Teaching a computer what to attend to is another open research problem. In people, attention is influenced by emotion—for example anger can focus attention on the object of the anger. Such cognitive-affective interactions are included in the fifth component of an emotion system.

The Roseman model is appealing for its simplicity and grounding in studies of human appraisals.[3] One limitation is that it does not address complex situations where multiple appraisals may be made. For example, if John thought that his teacher had designed an unfair test *and* that he himself was not prepared for the exam, then there would be two separate agencies, and it is unclear what he would feel—perhaps a mixture of anger and guilt. Nonetheless, the model provides a structure that could potentially be adapted

to this case. Overall, it shows promise for implementation in a computer, for both reasoning about emotion generation, and for generating emotions based on cognitive appraisals.

Emotion Synthesis via Multiple Mechanisms

The OCC and Roseman theories provide a rule-based mechanism for cognitive generation of emotions. The three examples I showed adapted the OCC model so that it could not only be used to reason about emotions, but also to synthesize affective states, to provoke emotional expressions, and in some cases, to prompt influences on a character's behavior, perception, and subsequent cognitions. The mechanisms used for all of this were relatively high-level, involving rule-based and case-based reasoning.

In humans, emotions are generated not only by explicit reasoning, but also by low-level noncognitive influences. We loosely referred to these as "physical" aspects early in the book, since they tend to be more easily associated with bodily phenomena than with mental phenomena. These aspects may only map metaphorically into non-embodied agents, but they are nonetheless relevant for mobile robots and other autonomous characters that at least simulate physical interactions with their environments. This section describes three models which encompass not only cognitive reasoning for generating emotion, but also additional low-level mechanisms, inspired by the human emotion system.

Four Elicitors for Emotion Synthesis

Carroll Izard (Izard, 1993) proposed that there are four types of elicitors of emotion in humans. These have inspired a new connectionist model of emotion synthesis, "Cathexis," developed by Juan Velásquez of MIT (Velásquez, 1996). The four elicitors in this model are:

- *Neural.* Effect of neurotransmitter and other neurochemical processes. These processes run independently, in the background, and are influenced by hormones, sleep, diet, depression medication, etc.
- *Sensorimotor.* Effect of posture, facial expression, muscular tension, and other central efferent activity. These effects primarily intensify a given emotional state, but in some cases appear to be capable of generating new affective states.
- *Motivational.* Effect of sensory provocations such as anger provoked by pain, of drives such as hunger, and emotions evoking each other.
- *Cognitive.* Effect of cortical reasoning, implemented here via an adaptation of Roseman's theory.

Cathexis consists of a constellation of proto-specialists, like Minsky's agents in the Society of Mind (Minsky, 1985). Each proto-specialist represents a basic emotion type, which receives inputs from the four elicitors, as well as from other proto-specialists. Each proto-specialist can exert influence on output behaviors, for example, joy, with intensity above its activation threshold, can produce a smile. Each can exert influence on other proto-specialists, for example, joy can inhibit distress, and activate hope. Since proto-specialists are used to implement both emotional and non-emotional states, it is easy for emotions to interact with physical states; for example, sorrow increases fatigue and decreases hunger. The result is a distributed connectionist-flavor model that can synthesize a variety of emotions simultaneously.

In contrast to the OCC model, where the structure of the rules varies for each emotion, the Cathexis model has only one update rule. The rule contains terms that take on values specific to proto-specialists, but otherwise the form is the same for every proto-specialist's emotion intensity. At each time t, each proto-specialist $p = 1 \ldots P$ updates its emotional intensity $I_p(t)$ as follows. Let $\varepsilon_{p,i}$, $i = 1, 2, 3, 4$ be the values contributed to proto-specialist p by the four elicitors.[4] Let $\alpha_{p,m}$ be the excitatory gain applied by proto-specialist m to proto-specialist p. Let $\beta_{p,m}$ be the inhibitory gain applied by proto-specialist m to proto-specialist p. Finally, let f be a function that controls the temporal decay of an emotion intensity, and let g be a function that constrains the emotion intensity to lie between zero and its saturation value. The new intensity is then a function of its decayed previous value, its elicitors, and influences from other emotion intensities:

$$I_p(t) = g\left(f(I_p(t-1)) + \sum_{l=1}^{4} \epsilon_{p,l} + \sum_{m=1}^{P} (\alpha_{p,m} - \beta_{p,m}) I_m(t)\right).$$

As in the OCC model, the intensity is compared to an emotion-specific activation threshold before determining if an emotion exists. Only if the intensity exceeds the activation threshold does the proto-specialist release its value to influence the behavior system and other proto-specialists. In addition, each proto-specialist has a saturation threshold. When the intensity exceeds this threshold, then it stops increasing. Mechanisms such as this contribute to the nonlinear behavior of this model. Temperaments are encoded in Cathexis via these thresholds, the parameters α and β, and the decay rate chosen for f. For example, an excitable temperament would be modeled as having lower activation thresholds; it would take smaller levels of stimuli to activate its emotions.

Emotion intensities above a certain threshold are allowed to influence a "behavior system," which is responsible for both emotional behavior and

emotional experience. The behavior system consists of a network of behaviors such as "make a fearful facial expression," and "run away." Each behavior consists of two components: an expression (e.g., smile) and an experience (e.g., feel happier). The experience implemented here can be thought of as the first aspect of emotional experience only; it does not implement sensations like human feelings. Emotional behaviors compete for control, with the value of each behavior determined by a linear combination of "releasing mechanisms," an ethological concept that includes internal motivations and drives, emotions, and external stimuli such as presence of friend or foe. For example, a combination of "anger" and "foe present" might release the "bite person" behavior.

Velásquez has implemented the Cathexis model in a scenario with "Simon the toddler." Simon is a synthetic agent representing a human toddler. Simon has proto-specialists for six basic emotions: fear, anger, sadness, happiness, disgust, and surprise, and for five drives: hunger, thirst, fatigue (need to rest and sleep), interest (need to explore and play), and temperature regulation. Different thresholds, excitatory and inhibitory gains, and decay rates can be chosen for each emotion to customize Simon's temperament. Cathexis provides the first complete example of a computational system that incorporates at least an approximation of all the major types of mechanisms known to be involved in human emotion synthesis. It is an important first step toward development of a complete computational emotion system.

A Three-Layer Architecture

Aaron Sloman, a philosopher at the University of Birmingham in the U.K., was one of the first to write to the computer science community about computers having emotions (Sloman, 1981). Sloman, assisted by students and colleagues, notably Luc Beaudoin, Ian Wright, and Brian Logan, has proposed and refined an architecture for human-like emotions. This architecture has not been thoroughly implemented and evaluated in computers; however, it has several features that make it relevant to affective computing, and especially to emotion synthesis.

Sloman conjectures that adult humans have at least three architectural layers in their brains: a reactive layer, a deliberative layer, and a self-monitoring layer. These three layers can be categorized loosely according to their evolutionary age—oldest to newest—and according to their functional similarity with other animals. An animal with just a reactive layer would have a tendency toward simple predictable behavior. For example, it might always run when it sees light, giving the impression of a "fear" behavior. In Sloman's architecture, the reactive layer detects things in its environment, and executes fairly automatic processes to determine how to react. Although the automatic

processes could in theory represent sophisticated behaviors, their speed and relatively "hard-wired" nature make them better suited for responses that need to be rapid and that rarely need to be modified. The reactive layer is capable of some simple learning; however, it is not able to construct or evaluate plans. Emotions such as startle and disgust are likely to be generated by this layer.

The deliberative layer is capable of planning, evaluating options, making decisions, and allocating resources. The emotions involved in goal-success or goal-failure, i.e., those which are cognitively assessed, are also found in this layer. This includes, for example, the poker-playing agent who is pleased at winning with a good hand. The deliberative layer is also capable of learning generalizations which, once reliably mastered, can be transferred to the reactive layer. Despite the flexibility of the deliberative layer, its performance can still be improved by a higher layer that monitors the long-term impact of its functioning.

The third layer, self-monitoring meta-management, prevents certain goals from interfering with each other, and can look for more efficient ways for the deliberative layer to operate, choose strategies, and allocate its resources. Sloman suggests that emotions associated with this layer might include shame, humiliation, and grief. In particular, use of this architecture for modeling grief has been explored (Wright, Sloman, and Beaudoin, 1995). One of the interesting phenomena that this architecture tries to explain is that of *perturbance*, whereby thoughts, previously rejected or postponed, resurface and interrupt your attention. For example, during grief, thoughts of the lost object of affection frequently perturb one's thinking. At the loss of a beloved friend, your thoughts are repeatedly interrupted to think about him or her.

The three-layer architecture is a potential model for emotion synthesis that compares favorably with findings in the neurological, psychological, and cognitive science communities. Its reactive layer would be where the "fast primary" emotions arise. These are the innate, hard-wired, or "compiled" processes, which execute without prior conscious cognitive appraisal. In humans and a variety of other animals, these functions reside in parts of the limbic system and lower brain stem. For example, disgust, as expressed on the face when something vile is placed on one's lips, occurs even in an infant born with only a brain stem, who does not survive long after birth. Moving up to the deliberative layer, we can make a correspondence with Damasio's so-called "secondary" emotions. These are the cognitively-generated emotions which typically require some kind of cortical reasoning about goals, situations, objects, and events. When either primary or secondary emotions arise, they can activate reactive processes which, in the human, would probably involve the amygdala, which subsequently activates bodily responses

comprising the physical aspects of an emotional experience. The third layer, meta-management, is the only layer where the notion of "self" is significant. Consequently, it is reasonable to hypothesize that it is where the "self-conscious" emotions such as shame, guilt, and embarrassment are likely to arise. These are the highly cognitive emotions, which appear to develop in childhood after the notion of self is intact. They are also more social, taking into account how people evaluate one another.

Although the three-layer architecture lacks details of implementation, it illustrates the need I have argued for multiple levels of models in emotion synthesis, including both low-level primary mechanisms and higher-level cognitive ones. In particular, it illustrates the need for a higher "self-monitoring" process for management of emotions. The latter is a crucial piece of a system if it is to develop the skills of emotional intelligence for regulating and wisely using its emotions.

Emotions, Hormones, and Homeostasis

Emotion synthesis raises questions about low-level "bodily" processes, because human emotions involve both the body and the mind. Even though computers do not have bodies like ours, they can simulate human bodily systems. Let's look at a model that explicitly simulates physiological changes relevant to emotion synthesis.

Dolores Cañamero, at the Free University of Brussels, has built a system in which emotions trigger changes in synthetic hormones, and in which emotions can arise as a result of simulated physiological changes (Cañamero, 1997). This system is part of a simulated two-dimensional world, with inhabitants called "Abbotts" and "Enemies." The Enemies do not have emotions in the present system, but some of the Abbott's behaviors, motivations and emotions are designed to deal with the Enemies. Each Abbott's behaviors, motivations, and emotions have corresponding physiological implications. In particular, the motivations are intended to be homeostatic. For example, when an Abbott walks around (behavior) its temperature increases, and when the Abbott is too warm (motivation) it seeks to decrease its temperature. Other Abbott behaviors include: attack, withdraw, drink, eat, play, and rest. Other Abbott motivations include: aggression, self-protection, thirst, hunger, curiosity, and fatigue. Each motivation has an intensity, and the one with the highest intensity gets to control both the Abbott's behavior, and what it attends to.

Motivation intensity, and therefore behaviors, are influenced by the Abbotts' emotions. Abbotts have six basic emotions: fear, anger, sadness, happiness, boredom, and interest. Emotions can be triggered by external events, or they can be triggered by internal physiological changes or patterns. For

example, fear is triggered if an enemy is present, resulting in increased heart rate and lower temperature. Alternatively, higher levels of endorphines can trigger a state of happiness. Emotions also influence perception; for example, a state of high endorphines reduces the perception of pain.

Cañamero's system illustrates the ability of a computer to simulate physiological elicitors of emotion, as well as emotion's influence on physiology. Such simulations, to the extent that they try to imitate human and other animal systems, are an important way to learn more about emotion synthesis and the influences of emotion. Additionally, we might make comparisons between functions of human physiological systems and functions of computer operating systems, such as the different ways in which both systems try to avoid intruding viruses, or the different ways in which both kinds of systems perform various regulatory functions.

Synthesizing Emotion's Influences

The focus in the previous section was on mechanisms for emotion synthesis, including both cognitive and non-cognitive elicitors. In this section I describe models for synthesizing emotion's interaction with other processes in the computer, specifically, how it can be used to realize multiple concerns, influence learning and behavior, and bias memory retrieval and decision-making. These interactions primarily address the fifth component of a system that has emotions.

Realizing Multiple Concerns

Human emotions play an important role in motivation and in helping people make decisions that realize their many concerns. Several researchers have suggested that emotions are manifestations of a system that realizes multiple concerns and operates with limited resources in an unpredictable environment. This principle is increasingly relevant for software agents and other computational devices that interact with people while trying to perform many tasks. Nico Frijda of Amsterdam University has set forth an appraisal theory of emotions based on this principle, which he describes in his book, *The Emotions* (Frijda, 1986). Let's look at an implementation of his theory to illustrate a way of building computer emotions to influence various regulatory processes.

Jaap Swagerman, a student of Frijda, has implemented a portion of Frijda's theory in the computer program ACRES, *Artificial Concern REalisation System*, (Frijda, 1987). ACRES' primary task is to handle knowledge about emotions while interacting with a user. It receives and accepts (or rejects) inputs from its user, such as the name of an emotion and its description. ACRES tries to

learn about what causes emotions to be generated by having a user present it with thousands of scenarios, imitating how humans acquire knowledge by vicarious experience. Throughout this interaction, ACRES keeps track of its own internal emotional state, and can show this to the user if the user asks to see it.

ACRES has multiple concerns, and periodically examines the state of affairs to assess if any of its concerns need addressing. For example, if it has not learned anything new for a while, then its "vicarious learning concern" may trigger a request to the user for more input, so that ACRES can improve its emotional knowledge. For example, ACRES might ask the user if a recent interaction was attractive or aversive to the user. Later, if ACRES is given a new scenario, it will try to guess which emotion that scenario would cause, based on its similarity to the scenarios ACRES has learned.

Here are six concerns ACRES tries to satisfy:

1. Avoid being killed.
2. Preserve reasonable waiting times, i.e. respond promptly.
3. Receive correct input.
4. Receive varied ("interesting") input.
5. Safety (preserving the concepts in ACRE's concept-based structure).
6. Vicarious learning (from the user's experiences).

Ideally, the system should continuously evaluate the relevance of all events for all six concerns, in parallel, even during task-oriented activity. This requires hardware to support parallel processing, and is only simulated in a truncated manner in ACRES.

ACRES has to decide which concern to execute. This is done by giving each concern an importance index, with "avoid being killed" having the highest index. This index is not the only factor in determining which concern gets precedence; the gravity of the situation is also assessed: "how many times has the operator repeated his instruction?" and "what is the status of the operator—how well has he treated me?" These change during the interaction, so that it is difficult to predict which of the multiple concerns will first reach above-threshold relevance. When a concern becomes active, then information processing capacity and memory are used for setting up and executing actions to further that concern.

When one of its concerns is active, ACRES can react emotionally. For example, if ACRES detects an agent that repeatedly threatens its safety, and that does not heed ACRES's requests to stop, then ACRES becomes angry at that agent and may restrict its access permissions. In general, ACRES

diagnoses the situation over time, generates an emotion, and chooses a meaningful action. In fact, "emotional," with its most juvenile connotations, is a suitable adjective for some of ACRES behavior. For example, ACRES will complain if the user types the wrong thing at it too many times. It can refuse to accept inputs if the user repeatedly mistreats it. It will also react with plaintive requests not to be killed if the user types "kill." ACRES' childish behaviors render it an unlikely prototype for the kind of affective computer any of us would want on our desk. It provides an example of a system that has emotions without having emotional intelligence.

Nonetheless, ACRES is an important testbed for exploring how emotions arise and influence behavior. In particular, ACRES demonstrates several important functions included in the fifth component of an emotion system: It uses emotion to juggle the demands of user requests with interrupting current tasks, with shifting resource allocation, and with initiating questions. Its use of emotions in these ways helps it realize multiple concerns and appraise relevance, potentially helping it choose more intelligent actions. ACRES illustrates that emotions are not just for entertainment, but that they can provide low-level regulatory functions needed by a system with limited resources and multiple goals operating in a complex and unpredictable environment.

Emotions Influencing Learning and Behavior

Emotions are hypothesized to provide the flexibility not present in traditional stimulus-response theories of learning. Mowrer and his colleagues, through many experiments, determined that learning is best thought of not as the single stage of stimulus-response, but as two stages, with the first involving the generation of an emotion (Mowrer, 1960). Consider a rat that learns to leave a box upon hearing a tone, after previously being presented with that tone paired with a painful shock. Mowrer's theory delineates two processes: (1) the rat learns *to fear* the tone and (2) the rat learns that leaving the box reduces his fear. The advantage of the two-process model is that it explains why, if a barrier prevents it from leaving the box, the rat will seek an alternate way to reduce its fear. The emotional state allows for more flexible learning, while simultaneously providing a source of motivation: fear drives the rat to explore methods of escape.

Implementing emotion's influence on learning is an important piece of implementing the fifth component of an emotion system. This piece can be illustrated by some of the work of Bruce Blumberg at the MIT Media Lab. Blumberg's animated dog, Silas T. Dog, does not have an explicit emotion model, emotional state, or mood, but its expressions and behavior are influenced by internal variables that represent emotions as well as other internal states such as hunger or thirst. Although Silas's emotions are simple and hard-

wired, Silas has a key feature that has yet to be incorporated in the other models: the ability to learn, and in particular for his emotions to influence what he learns. Changes in Silas's internal variables drive a learning process. For example, if Silas sees something that scares him—increases his internal variable of fear—then he tries to determine which stimuli from his perceptual inputs and short-term memory are the best predictors of the change. This enables Silas to learn the association between a fear-causing stimulus and the ensuing emotion. Thus, he can learn new ways to behave, such as avoiding a place where he previously saw something that scared him.

One of the problems with building creatures that exhibit emotions is how to map emotional states to behaviors. As we saw, fear motivates the rat to find a means of escape, but it does not automatically tell it what means to pursue. Emotions motivate and bias behavior, they do not completely determine it. Silas's internal variables provide a biasing mechanism for his behavior. The variables have global effects, biasing or predisposing him to certain behaviors or actions, without determining these behaviors or actions. A behavior is most strongly influenced by "releasing mechanisms," which recognize and signal the presence of an event, such as food being placed nearby, or a foe coming into the vicinity. A releasing mechanism that detects food will probably cause a hungry Silas to approach, but if he is feeling fearful he will approach differently than if he is feeling happy. The difference caused by the emotions is seen in his bodily movements and posture, such as how he holds his head. The releasing mechanism prompts a behavior, and the internal variables of emotion bias how the behavior is executed.

Silas is a creature with multiple goals, needs, and behaviors, but with limited resources for acting and fulfilling his needs. Emotion arises when Silas's goals are furthered or thwarted. For example, if his goal of playing succeeds he feels happy. When he is happy, he also will be more inclined to want to play. The introduction of new objects and events in his environment cause these feelings to change. For example, when a hamster enters his room, he will feel more aggressive and pick up the hamster to shake it. His aggression is programmed to decrease as he shakes the hamster, or if a human agent in the perceived environment signals Silas to do something else. These emotion-behavior links are mostly hard-wired in Silas; a general framework for how to enable such links is an open research area.

Computer scientists can already build machines that learn, at least in some ways, without explicitly giving them emotions; however, giving them emotions appears to be a means to achieve multiple goals, only one of which is more flexible learning. A single emotion accomplishes many things at the same time. For example, a negative emotion that produces a bad feeling may trigger reassessment of what caused the bad feeling, followed by learning how

to avoid it in the future. If, while learning, the machine predicts it will feel even worse if it does not forward you a piece of urgent and important news, then it might interrupt its own learning experience to get the news to you. Even negative emotions such as anger or frustration can be beneficial to a system—helping it focus on a goal, or triggering it to reassess a situation and look for a way to improve it. In other words, an emotional state produces internal control signals in a machine running several tasks at once, and can signal its attention when its time for a change. These same signals can influence not only learning, but also memory, perception, and many other important functions.

Affective Decision Making

One of the most intriguing influences of emotion in a human is on rational decision making. Flexible and intelligent decision making has been an elusive goal of AI researchers. Computer scientists have a trove of problems that exhibit combinatorial explosion—where one possibility opens up several new ones, each of which opens up several more new ones, and so forth. An efficient solution to such problems is the holy grail of computer science. On the other hand, humans solve intractable problems all the time, problems with an explosion of possible answers, where there is not time to evaluate them all. Furthermore, most human problems do not operate with a fixed numerable space of possibilities. In chess, the computer can describe which piece, if any, is at each of the 64 squares. Although, no computer can evaluate all the positions that could occur in the game, a number that is estimated to be greater than the number of atoms in the universe, a computer can at least characterize the space of such positions. In normal human situations, even the *space* of possibilities may change; the combinatorial explosion explodes again. Nonetheless, humans almost effortlessly make decisions that would stymie the world's fastest computers. Is it merely the case that we are that much better at pattern recognition, learning, and reasoning? There is no question we are better at certain tasks involving these tools, but I think that AI has ignored a crucial component that is even more basic to human problem-solving abilities: the use of feelings and intuition to guide reasoning and decision making. Let me suggest a model for emotion's influence on decision making by considering a scenario of a human making a personal decision.

Albert, a very busy scientist, has a beloved eight-week-old boy, and is trying to decide how to provide for his son while he works during the day. He does not know any family members or friends who could help. He acquires lists for three kinds of day care providers: a list of ten nanny-referral services, a list of 145 licensed family care providers, and a list of 24 day care centers located nearby. He contemplates posting notices in newspapers and on bulletin boards. Albert loves his son, and wants to choose

the best care for him. He needs a care-provider within a month. Albert is a highly rational man; how does he decide what to do?

Here is what Albert says. I have inserted [good] or [bad] to emphasize the valence of several of his statements:

I thought of posting notices but you hear of so many wierdos out there these days [bad] that I thought it would be safer first to try the three lists I got, since they include licensed and trained providers [good]. I decided to consider all three types of care equally, since I hear that their quality is largely a function of the people involved.

Nannies. I do not want to give up my privacy and have a live-in nanny, but it would be great to have one come to my home during the day as this would be the most convenient [good]. I know nannies are expensive, about twice the price of the other options [bad]. The nanny-referral services want huge fees up front before you find anyone [bad]. Nonetheless, I am willing to pay more if I could find an outstanding nanny who would be with us for many years [good]. I am concerned about finding a nanny in four weeks, as I just went through a stack of old newspapers, reading "nanny wanted" ads and saw the same ads for the last three weeks [bad]. I have also heard several stories about nannies lately. There was a television special about nannies who abused or neglected the kids during the day [extremely bad]; their behavior was observed on hidden cameras. However, this probably made the news because it is rare; abuse could happen with any care provider. I am also concerned about stability, which is important to a child. One of the guys at work told me they were on their third nanny this year [very bad]; they once again hired someone who said she would stay for at least a year, and then she changed her mind. Let me check the other options before I go further with this one.

Family care. What a huge list; this will take forever [bad]. I'll skip this for now.

Center care. Some friends recommended a place nearby that they love [good]. I visited and thought the place was too institutional [bad]. Center care appeals to me because it is stable, the people have training and are licensed [good], and there is usually someone to back them up so they can take breaks. This probably reduces stress and the chance of abuse [very good]. They do character checks on their employees and I can confirm their history with state offices to verify that there are no reports of abuse [good]. I started calling all the centers on the list that took infants. None of them had immediate openings [bad]. I made appointments to visit all the ones that indicated they might have an opening within a month. One of the places I really like and two others were pretty nice. I paid the fees to get on three waiting lists.

Back to the list for family care. I know people who are very happy with family care providers [good]. These can be stable and stimulating [good] but they may lack training and support when there are complications [bad]. I started calling everyone on the list. About one in twelve had an opening for an infant within a month [good]. I inquired how many kids they had, the environment, their experience, their assistants, what they do during the day, and the hours they worked. Some of the providers I ruled out on the phone; one of them sounded more interested in money than in children. I made appointments with the best sounding ones and started visiting. Several of the homes had huge dogs, one which looked like it could swallow my son. I added "no big dogs" to my checklist of criteria.

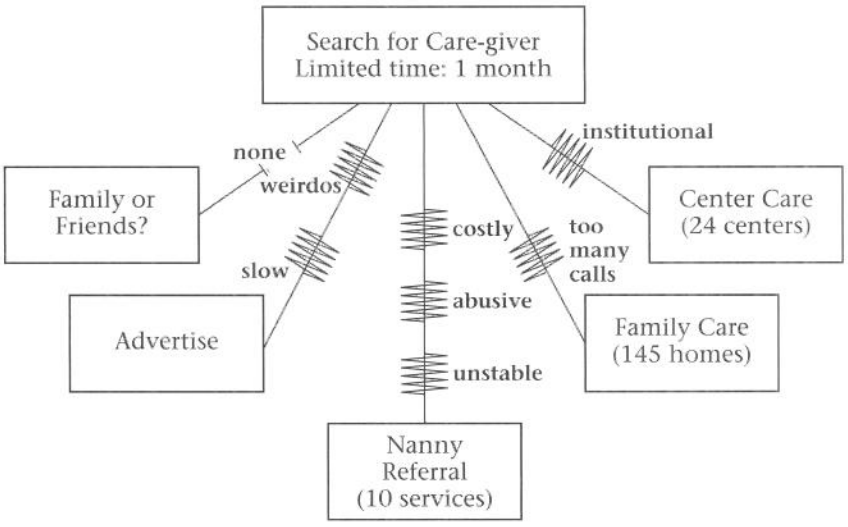

Figure 7.6
Initial consideration of child care options marks several possibilities with negative resistance due to bad associations.

Part of Albert's decision is illustrated in Fig. 7.6, where we see the five possibilities he considered, together with various "negative resistances" I have added to them, to model the valence associated with various pieces of the decision. For example, because he believed there were no family or friends who could help, this branch was effectively pruned off the tree of possibilities. Advertising met with two doses of negative resistance—fear of wierdos and fear that, like the ads he saw in the paper for so many weeks, his ad would go unfilled. The nanny option met with three pieces of negative resistance, and so forth. None of the negative resistances precluded further exploration of these options, but they biased him to first consider those associated with the fewest negative feelings.

Albert combined these valenced biases with many logical actions: he systematically gathered information about a variety of affordable options conveniently located in his town, ruled out those that did not take infants, ruled out those which were unavailable within the month, made appointments to meet people, visited the potential caregivers, learned about their environments and experience, and checked references. However, he did not logically find the best care, at least not in the sense of having weighed all the alternatives available to him. He did not generate all the possibilities, and did not have time to explore all the ones he did generate. Although the set of possibilities listed above looks manageable, in practice there were always others which could arise.

Furthermore, although he started with what he thought was his full list of criteria (full-time care, stability, etc.) this list did not forsee every criterion that would become important to him during the process, such as no big dogs. It was impossible to objectively state all the criteria up front; he discovered along the way what was most important to him. As he searched, he may have also learned of new possibilities—a great care-giver who could take his son four days a week; a wonderful neighbor who could help on the fifth day. In contrast, today's computers that conduct searches only guarantee optimal results if given precisely stated criteria, constraints, and a specific space to search.

Albert did not search all the possibilities, but he searched until he ran into either negative feelings or logical constraints, and then he stopped and tried something else. He continued this strategy—exploring possibilities that felt reasonable and good, and modifying his criteria as he accumulated more information. When he visited somebody and noted something new that resulted in a bad feeling, like the big dog, he added this to his criteria of things to avoid, crossed the site off his list, and continued with his search. Before his time ran out, he arrived at a decision that combined logical constraints with weighing the good and bad valences associated with his options. Emotions played an integral part not only in his final decision, but also in his process of gathering information.

To date, there are no computers with emotions that influence their decision making and other cognitive processes to the same degree that these influences are believed to occur in people. Nonetheless, computers could be given these abilities, especially when facing problems where the options cannot be fully explored. On the other hand, computers should not try to use affect for all decisions. There remain many problems involving possibilities that can be enumerated, where time permits a purely logical approach or a brute force search for finding the optimal solution. In such cases, computers are likely to be faster at finding the solutions than humans. There is no need to involve emotions in these kinds of problems, unless perhaps to contribute a positive feeling when the answer is found, which might reinforce the learning of that answer, if such a goal is desired.

Nevertheless, when a system faces problems where the possibilities cannot be enumerated and evaluated in the available time, I suggest that affective decision making provides a good solution. Humans use feelings to help them navigate the oceans of inquiry, to make decisions in the face of combinatorial complexity. These feelings might be called "intuition" or "a sense of knowing" or just "gut feelings." Regardless of what they are called, they provide a mechanism through which emotion works powerful influences on human cognition and behavior. People respond with remarkable intelligence and

flexibility despite insufficient knowledge, limited memory, and relatively slow processing speed. An integral component of human decision making is emotion, and this component could potentially be given to computers.

Emotions that Interact with Memory

The same emotion that influences a person's learning and decision making also influences memory retrieval and a host of other cognitive processes. Scientists believe that emotional valence attaches to concepts, ideas, plans, and every experience stored in our memories. Good feelings likely encode knowledge of effectiveness, familiarity, opportunity, and associations with positive outcomes. Bad feelings likely encode knowledge of ineffectiveness, unfamiliarity, risk, and associations with bad outcomes. When it is time to make a decision, valenced feelings help bias a person away from bad outcomes, and toward good ones. As studies of patients with prefrontal brain damage show, these biasing mechanisms are apparently at work *before* declarative knowledge for reasoning is activated. Furthermore, without the help of these mechanisms, the person may not be able to behave in an advantageous way (Bechara, Damasio, Tranel and Damasio, 1997). In other words, these biasing mechanisms occur both before and during pattern recognition and reasoning, greatly influencing their effectiveness. Let us consider how such mechanisms might be constructed for computers.

Because memory is intricately involved in decision-making and almost every aspect of cognition, it may be that the way in which emotion works so many of its influences is via its influence on memory. The findings of Bower and Cohen (1982) on mood-congruent memory retrieval and learning have influenced several models for representing emotion-memory interactions and their impact on cognitive processes.[5] These include the FEELER model of Pfeifer and Nicholas (1985), and the DAYDREAMER model of Dyer and Mueller (Dyer, 1987; Mueller, 1990). The latter is not only able to perform reasoning about emotions, but it also uses the appraisal process to generate an internal emotional state that influences the system's planning, learning, recall, and production of hypothetical scenarios, or daydreams, exploiting the influences of mood-congruent memory retrieval.

Let's take a closer look at a model inspired both by the findings on mood-congruent memory retrieval, and by the findings of LeDoux about the role of the amygdala and other sub-cortical structures in processing emotions. Aluizio Araujo, of the University of São Paulo in Brazil (Araujo, 1994) has built a model that attempts to integrate both low-level physiological emotional responses and their high-level influences on cognition. Araujo's model represents emotions via the dimensions of arousal and valence. It consists of two interacting neural networks—the "emotional network" and the "cogni-

tive network." These are designed to roughly approximate the roles of the limbic and the cortical structures in the human brain, respectively. The first network evaluates the affective connotation of incoming stimuli and outputs the emotional state of an individual. It performs relatively simple processing, providing a fast response, like limbic structures. The second network performs cognitive tasks such as free recall of words and associations of pairs of words. It performs more detailed processing on the inputs but provides a slower response, like the cortex.

Araujo's two-network model is designed to imitate mood-congruent memory retrieval and learning effects and the influence of anxiety and task difficulty on memory performance. He lists more than forty specific requirements of these interactions that his system attempts to satisfy (see Araujo (1994), pages 46–50). The essential aspects of his system are as follows: An "emotional processor" calculates arousal and valence for every stimulus. The arousal and valence produced by the emotional net influences cognitive processes by changing parameters on the cognitive net. Araujo acknowledges that both nets should influence each other to mimic the influences in the human brain, even though he only takes time to address the influence in one direction. In particular, the emotional net's outputs can influence the learning rate and accuracy of the cognitive net, influencing its performance, as well as what it learns and can retrieve. The model imitates anxiety's influence on learning (Spence and Spence, 1966). Araujo's model is a significant step toward combining emotion with memory, not as an "add-on" function, but as a closely intertwined mechanism.

Nonetheless, Arajuo's model does not solve a couple of the fundamental problems with implementing mood-congruent memory retrieval in computers. The first problem is that computers do not automatically have valence attached to everything they learn; some mechanism must determine if the item is good or bad. I described this "bootstrapping" problem briefly in Chapter 2 where I suggested that computers with bodies could have hardwired notions for bad—such as things that cause pain, and for good—such as things that relieve distress. These could be augmented with the ability to learn valence by association. However, computers without bodies will need to be given some other reference points for making judgments about valence. What these references should be is a significant open question, as they will largely bias what it learns as good and bad, and then right and wrong. This problem can be expected to raise questions in religion and ethics as well as in computer science.

The second problem raised by mood-congruent memory retrieval relates to the mechanisms for its implementation. How is valence encoded? Scientists believe that in humans, feelings encode valence—the same subjective

feelings that I described as the least understood part of human emotional experience. To give a computer feelings raises the problems I described earlier of consciousness and physiological sensing, for which scientists have yet to propose working solutions. One partial solution is to construct in a computer an extra bit for every item in memory, to carry valence information. This could be augmented with another bit or two for a coarse intensity value, and with dedicated parallel mechanisms for rapidly and automatically summarizing the valence of memories associated with a particular thought, so that this information is always available, even before being consciously requested. This solution imitates the behavior of human feelings in representing valence for stored memory items, and in providing a background process, akin to subconscious processing, for assessing valence. However, this solution still does not account for the complexity of emotion's interactions with memory in humans.

There is presently no satisfactory model for representing the mechanisms of feeling for signifying when someone knows something, or similarly, for handling the ability of certain stimuli to trigger a special feeling of meaning. These kinds of feelings are enigmas for the present; all people have them, but scientists do not understand them well enough for us to determine how to implement them in computers. Simulating physical systems, as in the hormone simulation above, is not the same as having awareness of physical sensations. Suppose that a computer had separate physical mechanisms for simulating physiological responses, even if not the same kind of responses as in the human body. Sensors in each system could receive biochemical and bioelectrical information from around the "body," which could then be communicated to a "conscious" unit to provide an awareness of bodily sensations. However, the nature of this awareness would still be quite different from that of a human, owing at least in part to the different physiology. Computers and humans have very different bodies, so computer and human feelings are likely to be very different. In other words, the emotional experience we can give to a computer does not duplicate that of humans; computers cannot feel what we feel. But, for that matter, we cannot verify that our own children can feel what we feel; we only guess that our similar physiology permits similar experiences.

Emotion as an Umbrella

There is a temptation to think of emotion as a single concept, and to try to define it with one all-encompassing definition. However, as we have now seen, the human emotion system consists of many different components, and not all emotions use all the components in the same way. To implement low-level fear, we need crude pattern recognition and fast responses, capable of hijacking other cognitive processes. To implement hope, we need

cognitive processes, with something that can be hoped for. For each emotion, researchers should detail the components it includes and determine which mechanisms are best for its implementation. My explanation of why the word "emotion" is defined in so many ways by different theorists is because it consists of many distinct components. Two different emotions may or may not share the same components. Consequently, when it comes to synthesizing emotion, different components are likely to require different mechanisms, like in the human brain, where we know fear blazes its own fast path through the limbic system, while emotions like hope are believed to be more cortical.

The term "emotion" is perhaps best thought of as an umbrella, under which a variety of processes cluster. When synthesizing emotion, therefore, we do not need to pick just one aspect of a cognitive, physiological, or behavioral model, but we need to consider how each of these works with the others. If the component is low-level, then signal-based representations and connectionist interactions may suffice for its implementation. To imitate certain bodily influences it may be necessary to construct biophysical models, as is being done in low-level modeling of fear (Armony, Servan-Schreiber, Cohen, and LeDoux, 1997). If an emotion is high-level, then rule-based reasoning may play a role. In either case, regulatory mechanisms will need to be a part of a complete emotion system. It is perhaps not best to try to build one rule-based model that makes all emotions, or one connectionist model that makes all emotions, and so forth. Instead, different models can be used for different mechanisms, and their interactions tailored in accord with the distinct nature of each emotion. This is not to say that a different model is necessary for every emotion; that much differentiation would lead to duplication of many components of each model. However, neither is one unified model the best solution. The answer lies in-between these two extremes. Once a suitable set of mechanisms are found, it is important to combine them all in the same system, to gather them under the same umbrella, to ensure that they can function cooperatively. At this point, the regulatory effects of emotion will truly be put to the test.

Summary

This chapter has described emotion synthesis, specifically focusing on models that employ both cognitive and non-cognitive mechanisms for generating emotion. Cognitively generated emotions have been the easiest to implement in AI systems, as emotion theories are usually described with rules and lend themselves directly to rule-based implementation in a computer. This chapter has also described various ways to synthesize emotion's influence on other processes: both cognitive and physical. The former has focused

on learning, decision making, and memory, while the latter has considered various regulatory mechanisms in a computer, as well as simulations of human physical systems.

These last three chapters have revealed a variety of tools—from low-level numerical representations of signals and patterns, to high-level rule-based representations of goals, preferences, situations, and the emotions to which they give rise. An affective computer can be expected to employ many levels of tools—combining both low-level and high-level models, using both numeric and symbolic representations, employing tools from signal processing, pattern recognition, learning, common sense reasoning, and more. Many of the pieces are in place, but there are no complete emotional systems to date, at least not any that rival those in humans.