This is a preliminary version from P. Thagard, *Mind: Introduction to Cognitive Science*, second edition, to be published by MIT Press in February, 2005. For the references, see http://cogsci.uwaterloo.ca/Bibliographies/cogsci.bib.html.

Chapter 10

Emotions

How do you feel about the following items? Group A: death, cancer, poison, traffic tickets, insults, vomit. Group B: lottery winnings, fine restaurants, sex, victories, babies, parties. For most people, the things in group A are associated with negative emotions such as sadness, fear, and anger, whereas the things in group B are usually associated with happiness and pleasure. If you think of the main events of your day so far, you will probably be able to recall the emotions that accompanied them, for example the joy you felt when your sports team won, or the worry you felt when you realized that exams are coming soon.

Traditionally, cognitive science ignored the study of emotions, seeing it as a side issue to the more central study of cognition. Philosophers as far back as Plato have tended to view emotion as a distraction or impediment to effective thought. In the past decade, however, there has been a dramatic increase in the appreciation of the relevance of emotion to cognition, particularly with respect to decision making. On the old view, decisions can be made either rationally or emotionally, and cognitive science is mainly concerned with the rational ones. In contrast, the view is now emerging that emotions are an inherent part of even rational decision making. This chapter will describe how emotions contribute to both representation and computation. But first we need to discuss the nature of emotion.

WHAT ARE EMOTIONS?

Everyone is familiar with basic emotions such as happiness, sadness, fear, anger, disgust, and surprise. But there is much disagreement among cognitive scientists concerning the nature of emotions. Favored theories fall into two general camps, one viewing judgments largely as judgments about a person's general state and the other emphasizing instead bodily reactions. Suppose you were driving a car this morning and another driver suddenly cut in front of you and made you veer rapidly to the side of the road. Probably your first reaction was fear, followed by anger at the driver who cut you off. According to the view of emotions as judgments, your fear consisted primarily of an inference that you were at risk of bodily harm, violating your goal of staying alive and healthy. Similarly, your anger consisted of a judgment that the other driver was responsible for putting you in danger. Proponents of the view that emotions are primarily judgments include Oatley (1992), Nussbaum (2001), and Scherer, Schorr, and Johnstone (2001).

Oatley (1992) describes how basic human emotions are intimately connected with goal accomplishment. People are happy when their goals are being accomplished, and sad when they are not. If you do well on an exam or in a job interview, or if you get invited to a party, happiness derives from this satisfaction of your professional and social goals. Failure to satisfy such goals can produce disappointment and sadness. People

become angry at whatever frustrates their goals, for example someone stealing your parking space. You experience fear when your survival goals are threatened, as when a truck comes skidding toward your car. Disgust reflects a violation of your eating goals, as when someone offers you a chocolate-covered cockroach to eat. We can therefore see emotions as involving a very general representation of a person's overall problem-solving situation.

Why should such a general representation be used? Why not simply have verbal or visual representations that display the current status of goal accomplishment? Oatley points out that human problem solving is often very complex, in that it involves multiple conflicting goals to be accomplished, rapidly changing environments, and rich social interactions. Emotions provide a summary *appraisal* of your problem-solving situation that makes two important contributions to subsequent thinking. Appraisal that certain aspects of your situation are extremely important to your goals can lead you to *focus* on those aspects, concentrating your limited cognitive resources on what matters. Moreover, emotions provide readiness for *action*, ensuring that you will be spurred to deal with your problem solving situation rather than being lost in thought (Frijda, 1986). Thus emotions are not just incidental, annoying features of human thought, but have important cognitive functions concerned with appraisal, focus, and action.

Because emotions play a role in human thought and action, explaining why people do what they do often requires us to refer to emotional states. "He slammed his fist into the wall because he was angry." "She was smiling all day because she was ecstatic at being admitted to medical school." Sometimes, emotion-based explanation goes beyond verbal representation when we understand other people's emotions by imagining ourselves in their situation and experiencing an emotion that approximates to what they feel. This kind of understanding is called *empathy*. It is based on analogical thinking where you develop a mapping between someone else's situation and your own that actually produces in you some image of the emotion that the other person is experiencing (Barnes and Thagard, 1997).

In contrast to the view of emotions as appraisals, the opposing view emphasizes bodily reactions rather than cognitive judgments. When a driver cuts you off, you probably experience physiological changes such as increases in your heart beat, breathing rate, and blood pressure. On the physiological view, your fear and anger consist of your brain reacting to these physiological changes rather than making a judgment about your general situation. The view that emotions are largely a matter of physiological reactions originated with William James (1884). Damasio (1994) refers to the signals that the body sends to the brain concerning physiological factors as *somatic markers*.

There is no need to choose between a cognitive theory of emotions as judgments and a physiological theory of emotions as neurological reactions to bodily changes. Morris (2002) uses recent discoveries about neural functioning to argue that emotions depend on *interaction* between bodily signals and cognitive appraisals. What needs to be developed is a neurocomputational theory that shows how your emotions can involve both judgments about how the current situation is affecting your goals and neurological assessments of your body's reaction to that situation. I will describe such a theory below in the section on emotional computation, and extend it to consciousness in chapter 11.

Emotions such as fear, anger, and happiness involve reactions to particular situations: you are happy that you got a good grade on an exam, or angry that a friend failed to meet you as planned. In contrast, moods are much more long lasting and less directed toward particular situations. You can be in a good or bad mood for hours, without there being a particular thing or event that you are in a good or bad mood about. Psychologists use the term *affect* (with the emphasis on the first syllable) to encompass emotion, mood, and sometimes also motivation.

REPRESENTING EMOTIONS

What do emotions add to your ability to represent the world? From the representational theories described in chapters 2-9, it might seem that there is no need for a person or robot to have any mental structures beyond propositions, concepts, rules, analogies, images, and distributed representations. But tying these kinds of mental structures to emotions provides an efficient way to guide action. If you react to the prospect of eating sheep brains with disgust, then you do not need to do a lot of inference in order to decide what to do when you are offered some. In contrast, if your emotional association with ice cream is highly positive, then you will be strongly inclined to eat it. As my group A and B examples at the beginning of this chapter showed, we have positive and negative associations with many concepts.

Fazio (2001) reviews a large body of psychological experiments concerning the automatic activation of emotional attitudes attached to concepts. In these experiments, participants are given a word such as "cockroach" intended to prime a negative or positive attitude. They are then given an evaluative word such as "disgusting" and are asked to indicate as quickly as possible whether the word meant "good" or "bad". Many experiments have found that people are faster to answer when the prime concept fits emotionally with the evaluative word. For example, "cockroach" followed by "disgusting" produces a quicker response to "bad" than does "chocolate" followed by "disgusting". It therefore seems that emotional evaluations are closely tied in with the representation of concepts and objects.

People also have emotional associations with many propositions, analogs, and images. Depending on your country and interest in sports, the proposition that Canada won the 2002 Olympic ice hockey championship might be associated with excitement, disappointment, or boredom. Similarly, your attitude toward different rules that you have learned is probably represented not by some abstract numerical value like strength but by emotions that you associate with them. For example, the rule *IF you avoid early morning classes, THEN you can sleep more* is probably associated with happiness and relief. Analogs describe more complex situations than simple propositions and rules, but also can be associated with emotions. If you remember taking a course in which boring material and a demanding professor caused you to get a bad grade, then your memory of the course is probably associated with emotions such as disappointment and even anger. Any analogous course will prompt similar emotions that will steer you away from taking it. Images can also have positive and negative emotions: contrast your reaction to a hideous face from a horror movie with your reaction to the smiling face of your favorite movie star.

Different emotions can be distinguished with respect to two dimensions: pleasure and arousal. Figure 10.1 charts some important emotions with respect to these two dimensions. Where would you place yourself on this wheel right now? Presumably, both

the pleasant/unpleasant dimension and the aroused/unaroused dimension correlate with bodily reactions, and high arousal corresponds to more extreme physiological changes. But the dimensions should also correlate with cognitive appraisal of the situation that prompts the emotion, as when anger depends on realization that someone is thwarting your goals.

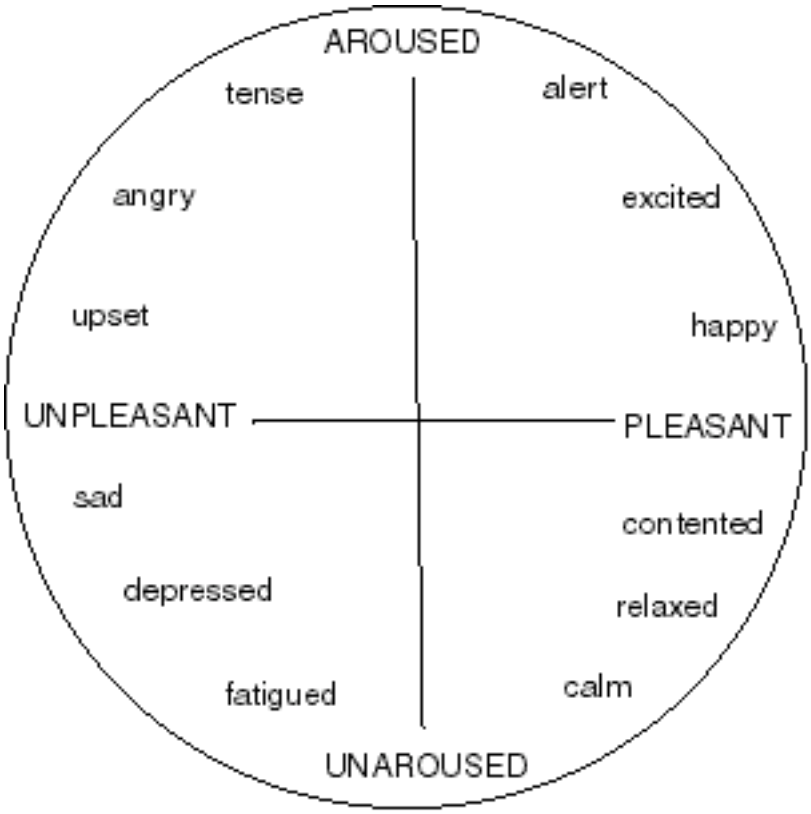

Figure 10.1. The structure of emotions with respect to pleasantness and intensity.

It is clear that emotions are represented in the brain, but it is much harder to say *how* they are represented. We have verbal concepts such as *happy, sad,* and *angry,* but there is a lot more to emotional thinking than just a connection between these concepts and other representations. A local neural network representation like the ones described in chapter 7 could have an excitatory link between a unit representing *ice cream* and a unit such as *happy*, but treating an emotion as just another concept conceals its links with judgment, physiology, and feeling. More plausibly, we should think of an emotion as a representation involving a pattern of activation across many neurons, as in the distributed representations discussed in chapters 7 and 9. Better yet, the neurons across which emotions are distributed should have connections to many different brain areas, including ones involved in cognitive judgments such as the prefrontal cortex and ones that receive inputs from bodily states such as the amygdala. From this perspective, an emotion is a pattern of activation in a population of neurons with connections to both inferential and sensory brain areas. The next section will describe a computational model of how this might work.

COMPUTING EMOTIONS

I once overheard the following conversation between two leading cognitive scientists. "Tell me, Maggie, could a machine have emotions?" "Well, Gordon, I'm a machine, and I have emotions." This response seems odd, because we do not usually think of people as machines or think of machines as having emotions. There are several positions that need to be considered:

1. Emotions have nothing to do with computation.
2. Computers can be used to model emotional processing in the brain, but emotion is not really computational.
3. Emotions can be a general function of computational intelligence, so they can arise in any sufficiently complex computer or robot.
4. Emotions arise from the particular kinds of computation performed by the brain.

There is not yet conclusive evidence for any of these positions, but I think that the last one is the most plausible. Hence the computational-representational understanding of mind should use be expanded using neurological findings to encompass emotions.

The simplest way to introduce emotions into a computational model is to add emotion nodes to the kind of local connectionist network described in chapter 7. Nerb and Spada (2001) present a computational account of how media information about environmental problems influences cognition and emotion. When people hear about an environmental accident, they may respond with a variety of emotions such as sadness and anger. Following the appraisal theory of emotions, Nerb and Spada hypothesized that a negative event will lead to sadness if it is caused by situational forces outside of anyone's control. But an environmental accident will lead to anger if someone is responsible for the negative event. If people see themselves as responsible for the negative event, then they feel shame; but if people see themselves as responsible for a positive event, they feel pride. Nerb and Spada (2001) show how determinants of responsibility such as agency, controllability of the cause, motive of the agent, and knowledge about possible negative consequences can be incorporated into a coherence network called ITERA (Intuitive Thinking in Environmental Risk Appraisal).

ITERA is an extension of the impression-formation model of Kunda and Thagard (1996). The main innovation in ITERA is the addition of units corresponding to emotions such as anger and sadness, as shown in figure 10.2. ITERA is given input concerning whether or not an accident was observed to involve damage, human agency, controllability, and other factors. It then predicts a reaction of sadness or anger depending on their overall coherence with the observed features of the accident. This reaction can be thought of as a kind of emotional summary of all the available information.

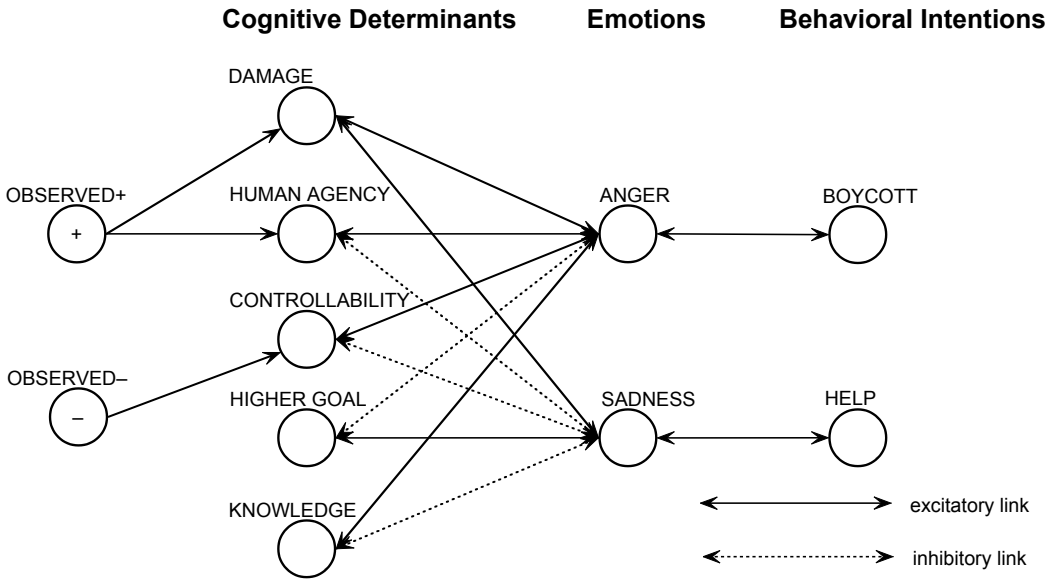

Figure 10.2. ITERA network for emotional reactions to environmental accidents. Solid lines are excitatory links, and dashed lines are inhibitory links. This example represents a situation in which a media report states that there is damage caused by human agency that could not have been controlled. From Nerb and Spada (2001), p. 528.

Another way to introduce emotion into local neural networks is found in the HOTCO (hot coherence) model of Thagard (2000, 2003). This model differs from other neural network models in that each unit is given a valence, representing its emotional value, in addition to the normal activation, representing its applicability to the current situation. For example the proposition that you have an exam tomorrow might get a high activation representing its truth and a negative valence representing how you feel about it. Valences spread through a network much like activations providing an overall emotional assessment of a thing, concept, or situation.

The ITERA and HOTCO models are unrealistic neurologically in the key respects discussed in chapters 7 and 9: they use local units that are not like real neurons, and they neglect the division of the brain into functional areas. Both these limitations are overcome in the GAGE neurocomputational model of emotional decision making (Wagar and Thagard, forthcoming). GAGE is named in honor of Phineas Gage, a nineteenth-century railroad worker who suffered brain damage because of a pipe that penetrated his skull. Amazingly, he survived his injury and regained his verbal and mathematical abilities, but became incapable of making sensible decisions about his work and personal life. Damasio (1994) describes modern patients with similar problems, and hypothesizes that they have lost the ability to make effective decisions because of disruptions in connections between the emotional and cognitive parts of their brain.

The particular brain area affected in Phineas Gage and similar patients is called the ventromedial (bottom-middle) prefrontal cortex, which provides links between areas of the cortex involved in judgments and areas involved in emotions and memory, the amygdala and hippocampus. Wagar and Thagard (forthcoming) showed how Gage's deficit and related psychological behavior can be modeled using groups of spiking neurons corresponding to each of the crucial brain areas: the ventromedial prefrontal cortex, the amygdala, and the nucleus accumbens, a region strongly associated with

rewards (it is also heavily involved in addiction to drugs and alcohol). The structure of this model is shown in figure 10.3. The GAGE model uses distributed representations of input stimuli and associated emotions, and relies on the spiking properties of neurons to provide temporal coordination of the activities of different brain areas. It is thus much more neurologically realistic than the ITERA or HOTCO models.

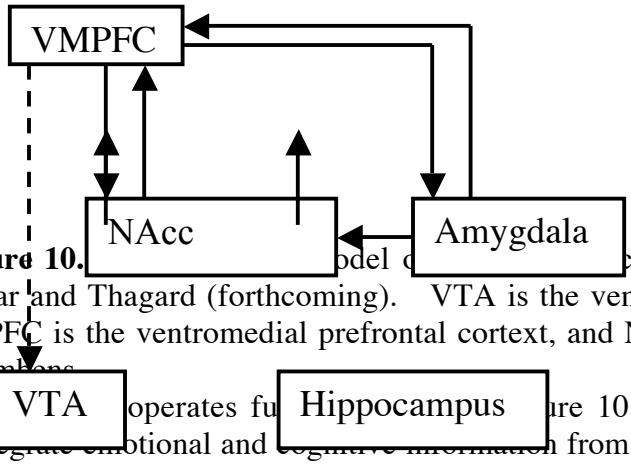

Figure 10. odel cision making, from Wagar and Thagard (forthcoming). VTA is the ventral tegmental area, VMPFC is the ventromedial prefrontal cortext, and NAcc is the nucleus accumbens.

Whe operates fu ure 10.4, the nucleus accumbens serves to integrate emotional and cognitive information from different parts of the brain. But when the model is "lesioned" by disrupting the neurons corresponding to the ventromedial prefrontal cortex, the model behaves like Phineas Gage and Damasio's patients, failing to integrate cognitive and emotional information. GAGE is not a full model of cognitive/emotional processing, but it shows how a such a model could take into account both cognitive aspects of judgment and appraisal performed by the prefrontal cortex and physiological input mediated by the amygdala. Notice that the connections in figure 10.4 among the ventromedial prefrontal cortex, nucleus accumbens, and amygdala are all bi-directional, suggesting that interaction is the right way of thinking about the relation between cognitive and physiological aspects of emotion. Thus the GAGE model shows how to integrate cognitive appraisal and somatic marker theories of emotions.

The three models described in this section are clearly just models: no one would claim that ITERA, HOTCO, or GAGE really have emotions. The problem is not just that they lack conscious experience of emotion (see the discussion of consciousness in chapter 11), but also that they clearly do not have the bodily inputs that are a crucial part of human emotions. The GAGE model takes for granted that the amygdala collects information about bodily states, but as a computer model it does not really have any bodily input. Because robot bodies are so different from human ones, I doubt that computers and robots will ever have emotions at all like humans: see the discussions of bodies in chapter 12 and the ethical implications of artificial intelligence in chapter 14. Nevertheless, it not implausible to describe the brain as a kind of emotional computer that integrates emotional and other sorts of information to enable people to make decisions. Psychological, neurological, and neurocomputational understanding of how emotions influence thinking is rapidly developing, so cognitive science is clearly responding well to the emotion challenge, in keeping with its response to the brain challenge.

What does the explosion of work on the neuroscience of emotions tell us about the philosophical positions in chapter 8 concerning the mind-body problem? As more and more is understood about the neurology of emotion and cognition, the dualist view that mind is sharply separate from brain becomes less and less credible. In addition, the functionalist view that mind is a purely computational construct independent of physical realization is becoming less plausible: human minds depend much more directly on human brains and bodies than functionalists would allow. Cognitive science is thus providing support for some version brain-based materialism, while still needing to deal with the problem of consciousness discussed in chapter 11. A theory of emotions is woefully incomplete if it cannot explain why we experience *feelings* such as happiness and sadness.

Functionalists could reply that as computer power continues to increase it should become possible to develop hardware and software that duplicate the structure and function of the brain to such an extent that computers do not just model emotions, they *have* them. There are already sophisticated computer models of thousands of neurons, so why not just expand them to billions of neurons organized into the relevant brain areas? Such models will be very useful, but they will inevitably simplify the extremely complex biological functioning of the brain. Currently, it is not technically feasible to model in full chemical and biological detail the operation of a single cell, let alone the complex of neurons with electrical and chemical signaling that comprise the brain. Therefore, although emotions can be simulated computationally and can plausibly be viewed as part of computational of the brain, it is unlikely that we will ever build a machine that duplicates human emotions. The ethical implications of this limitation are discussed in chapter 14.

PRACTICAL APPLICABILITY

In keeping with recent dramatic advances in experimental and theoretical research on emotion, there has been a sharp rise in investigations of the significance of emotion for many practical areas, including design, management, and the study of mental illness. Donald Norman, one of the most influential thinkers on the nature of design of computers and other artifacts, has a new book emphasizing the importance of emotion for design (Norman, in press). His slogan is: *Attractive things work better*. For example, he describes an experiment performed in both Japan and Israel comparing the ease of use of two forms of automated bank teller machines. Both forms were identical in function, but only one had the button and screens arranged attractively. People found that the attractive ones were easier to use, which suggests that beauty and function are interconnected. Norman argues that emotions change the ways that people solve problems, so designers should take emotions into account when they produce objects intended for human use.

Emotion is starting to be considered as an element in the design of intelligent computers. Picard (1997) advocates *affective computing,* which is computing that relates to, arises from, or deliberately influences emotions. She notes the importance of emotion in human communication, and argues for the value of giving a computer the ability to recognize, express, and respond intelligently to human emotions. For example, learning can be a highly emotional experience: think of times when you have been deadly bored by a lecture, or when you have been excited and inspired by a dynamic

teacher. We can expect a computer tutor to be much more effective if it can recognize the emotional state of the learners it is supposed to teach.

Understanding how emotions influence human thinking is also important for improving human decision making. Psychologists are increasingly appreciating that decision making is an emotional process as well as a cognitive one (Loewenstein et al., 2001). Finucane, Peters, and Slovic (in press) review evidence that judgments and decisions are influence by positive and negative evaluations attached to mental images of different situations. For example, affect-laden imagery contributes to preferences for investing in new companies in the stock market and adolescents' decisions to take part in health-threatening and health-enhancing behaviors such as smoking and exercise. This research is consistent with the view of Damasio discussed earlier that emotion is an inherent and ineliminable part of human decision making. Hence if you want to help people to make better decisions, it is crucial to take their emotions into account.

Understanding of emotions is also an essential component of professional success. Goleman (1995) advocates *emotional intelligence* as a complement to traditional notions of intelligence as purely cognitive. He describes how success in any social situation requires the ability to recognize and regulate one's own emotions as well as the emotions of others. Successful leaders need to be capable of empathy for others and able to inspire them by providing motivating emotions.

Finally, understanding of emotion is important for diagnosing and treating mental illnesses that involve emotional disturbances, such as schizophrenia and bipolar disorder (manic-depressive disorder). The causes of such disturbances are increasingly being identified at the molecular level, involving neurotransmitters such as dopamine and serotonin. Serotonin is a particularly important neurotransmitter that affects many kinds of behavior. Millions of people have been prescribed the drug Prozac for problems that range from depression to obsessive-compulsion to excessive anger (Kramer, 1993). Although side effects have been widespread, many people report substantial improvement in their emotional states as the result of Prozac therapy. Prozac increases serotonin uptake, making the neurotransmitter more readily available to particular neurons. Emotional changes such as falling in love involve neurotransmitters such as dopamine and hormones such as oxytocin. Hence anyone interested in increasing understanding of the role of emotions in human thought and mental illness has to pay attention to evolving knowledge about the mental effects of different neurochemicals. Brain anatomy is similarly relevant: Davidson et al. (2002) review research that identifies regions of the brains involved in depression, including the prefrontal cortex, the hippocampus and the amygdala.

SUMMARY

In contrast to the early decades of cognitive science, current research in psychology, neuroscience and even artificial intelligence is seriously concerned with emotions. There is increasing recognition that mental representations are often associated with emotional evaluations that contribute to many cognitive processes, especially decision making. Brain areas that support emotional processing include the amygdala and the prefrontal cortex. Computational models are being developed that show how decision making and problem solving integrate emotions with other kinds of information. Understanding of emotions is also contributing to practical applications such as design, education, management, and mental health. Appreciation of emotions

does not require abandonment of the computational-representational understanding of mind, but expands and supplements it in valuable ways that takes into account the detailed structure and functioning of the brain, right down to the molecular level of neurotransmitters. Emotion is highly relevant to the social nature of cognition, which is discussed in chapter 13.

DISCUSSION QUESTIONS

1. Can emotions be thought of as representations?
2. Do emotions get in the way of human thinking, or do they contribute to it? Would you want to have an emotionectomy?
3. What would it take to give a robot emotions?
4. How do emotions influence your education and work?

FURTHER READING

Research on emotion is highly diverse. For general reviews, see Davidson, Scherer, and Goldsmith (2003) and Lewis and Haviland-Jones (2000). Oatley, Catley, and Jenkins (1999) survey the psychology of emotions. On the neuroscience of emotions, see LeDoux (1996), Panksepp (1998) and Rolls (1999). Wierzbicka (1999) discusses emotion words in many languages. For philosophical discussion of emotions, see Griffiths (1997) and Nussbaum (2001).

WEB SITES

The emotion home page:
http://emotion.salk.edu/emotion.html
Affective computing at MIT:
http://affect.media.mit.edu/
University of Birmingham cognition and affect project:
http://www.cs.bham.ac.uk/~axs/cogaff.html
Geneva emotion research group:
http://www.unige.ch/fapse/emotion/
Kismet, a robot that models emotions:
http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html