Affective Computing

Rosalind W. Picard

The MIT Press

From The MIT Press

MITCogNet

First MIT Press paperback edition, 2000

This book was set in Stone Serif and Stone Sans by Windfall Software using ZzTEX and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Picard, Rosalind W.
Affective computing / Rosalind W. Picard.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-262-16170-1 (hc.: alk. paper)—978-0-262-66115-7 (pb.: alk. paper)
1. Human–computer interaction. 2. User interfaces (Computer systems). I. Title.
QA76.9.H85P53 1997
004'.01'9—dc21 97-33285
CIP

10 9 8 7

2 *Affective Computers*

Emotions are important in human intelligence, rational decision making, social interaction, perception, memory, learning, creativity, and more. They are necessary for intelligent day-to-day functioning. The negative connotations of "being emotional" or "acting emotionally" are not valid excuses for ignoring the study of emotions, or its application to computers. Instead, it is time to examine how emotions can be incorporated into models of intelligence, and particularly, into computers and their interactions with humans.

To date, researchers trying to create intelligent computers have focused on problem solving, reasoning, learning, perception, language, and other cognitive tasks considered essential to intelligence. Most of them have not been aware that emotion influences these functions in humans. Some have scoffed at the idea of giving computers emotions. However, now there is a preponderance of evidence that emotion plays a pivotal role in functions considered essential to intelligence. This new understanding about the role of emotion in humans indicates a need to rethink the role of emotion in computing.

Let me remind the reader, that when I refer to "computers" I mean not just a monitor and keyboard with one or more CPU's, but also computational agents such as software assistants and animated interactive creatures, robots, and a host of other forms of computing devices, including "wearables," which I will describe in greater detail later. Any computational system, in software or hardware, might be given affective abilities.

Most of today's computers do not have emotions *per se*. What would it mean for a computer to "have emotions"? To recognize or express emotions? To exhibit emotional intelligence? After a note about emotional development, this chapter considers these four topics in sequence: computers that recognize emotions, express emotions, have emotions, and have emotional intelligence. Along the way I will propose design criteria for such systems, describe tests they might have to pass, and examine the differences between emotion in computers and humans.

Developing Emotions

There is a tendency, perhaps because these examples are easy to think of, to imagine the worst forms of emotions in computers. For example, consider if you repeatedly typed the wrong input to your machine, and the machine finally said, "You stupid oaf! Read the manual." Clearly, there must be more to benefit from affective computing than this. This scenario, and similar ones involving grumpy agents or surly software, are products of a faulty underlying assumption—that computers with emotional states will also "act emotionally," which is virtually synonymous with "act stupidly." Having emotions does not *cause* stupid behavior. An intelligent adult has a full range of emotions and is capable of effectively managing and using these emotions to aid in many important functions. If he acts emotionally in a socially maladroit way, then he is thought to be less intelligent. Affective computers with poor emotional skills would be much worse to interact with than non-affective computers. If computers act stupidly when they first become affective, it will be because their designers have not given them the ability to act intelligently with emotions.

Consider emotional development in humans. Infants communicate primarily with emotional expression: crying, smiling, screaming, or laughing. All of these appear before language; emotions are evidently in the substrate before the more obvious signs of intelligence develop. If a child did not have the ability to express emotions, she would be severely impaired in her development and chances for survival. Emotional expressions are the signals used by an infant to communicate her needs: she cries when cold, bored, dirty, hungry, or in pain. A baby that smiles gives her parents the rewarding feedback of believing they have made their child happy. The infant does not encourage the parent by speaking, "Thank you for spending every other hour of the last twenty-four feeding me," but when the exhausted mother sees her child's contented expression, she *feels* a big reward, an emotional and intellectual boost that encourages her to continue caring for the child despite the effort involved.

Even though computers do not feel cold, hungry, wet, or bored in the same way a baby does, they have needs that might give rise to infant-like emotions. Every machine has an operating temperature range, and when the machine gets too hot or cold it fails. Computers need energy to run, and will cease processing if this is not provided. A robot's memory can fill up and "leak," making a mess which, although not offensive in odor, takes time to clean up. When given the goal of sorting through the overwhelming amount of networked sources of information, in search of interesting items for its user, then it might be useful for an agent to have boredom—to alert it to the need to stop wandering along a fruitless path.

As children mature, they learn social skills and ways in which to control their emotions and their emotional expression. As they develop, they also improve their ability to recognize emotions, to recognize situations that are apt to generate emotions, and to manage emotions. Similarly, an affective computer will probably need a developmental process whereby it acquires knowledge relevant to its affective and other abilities. It may be impossible for its designers, builders, and programmers to think of everything it will need in advance. Consequently, it will need both an innate set of abilities, and tools to continually learn new ones. The goal would be for it to reach an equivalent of "adulthood" in terms of affective abilities. A computer that interacts with adults should be capable of operating with the emotional intelligence of an adult.

Adult emotional intelligence consists of the abilities to recognize, express, and have emotions, coupled with the ability to regulate these emotions, harness them for constructive purposes, and skillfully handle the emotions of others. I use "emotional intelligence" in the way that has become common in the literature, even though there is ongoing debate about the use of the word "intelligence" since often the word implies something innate, whereas many of the aspects of emotional intelligence are skills that can be learned. Along the way it will become clear that every computer does not need all of these affective abilities all of the time. In fact, there will be many examples in the next chapter of applications where only a subset of the abilities are needed. Nevertheless, some computers that interact with people and have their own emotions could benefit from all of these abilities.

In the rest of this chapter, I propose criteria for giving computers affective abilities. These criteria are based on what is known about human emotions, on the methodology of adapting human emotions to computers. Of course, there are other possible methodologies that might yield different criteria; we are free to endow computers with any abilities that we can figure out how to develop. One of the most compelling alternate methodologies is that of letting emotions "emerge" in computers according to their own requirements. Since computers presently have different needs and behaviors than humans, why should they not be allowed to develop the emotions that suit these needs, as opposed to being given a set of our emotions that does not necessarily serve them well? This argument is a valid one, as long as the computers' needs remain subservient to the human needs for which the computer was designed. On the other hand, the word "computers" in the previous sentence can take on many different meanings and roles, one of which may include that of a social agent interacting with humans, in which case we can argue that the computer would at least benefit from understanding social aspects of human emotions. Consequently, even if social computers develop their own mechanisms of emotion, they will likely benefit from understanding human emotions, and end up with at least some affective abilities that are similar to human ones.

Adapting human emotions for computers should help computers acquire some of the benefits of emotions described in the previous chapters: more flexible and rational decision-making, the ability to address multiple concerns in an intelligent and efficient way, the ability to determine salience and valence, more human-like attention and perception, and numerous other interactions with cognitive and regulatory processes. Human-like abilities to recognize affect should also make it easier for computers to perceive human responses such as "pleased" or "displeased," which will help them learn how to adjust their behavior. This goal is motivated by a principle I would like to see practiced more: computers should be adapting to people rather than vice-versa. Facilitating the kind of interaction that comes naturally to humans is a win; it is a key step toward human-centered computing. Based on adapting what is known about human emotions, the next four sections propose, respectively, criteria for a computer to recognize emotions, to express emotions, to have emotions, and to have emotional intelligence.

Computers that Recognize Emotions

One of the hallmarks of an intelligent computer will be its ability to recognize emotions—to infer an emotional state from observations of emotional expressions and through reasoning about an emotion-generating situation. The computer might try to recognize the emotions of its user, of other agents with which it interacts, and of itself, if it has emotions. Recognition may require vision and hearing abilities for gathering facial expressions, gestures, and vocal intonation. Additionally, the computer may use other inputs that may or may not have analogs in human senses—reading infrared temperature, measuring electrodermal response, and so forth. Once emotional expressions are sensed and recognized, the system can use its knowledge about the situation and its knowledge about emotion generation to infer the underlying emotional state which most likely gave rise to the expressions. Giving a computer these perceptual and interpretive abilities could potentially give it as much ability to recognize emotions as another person might have. How will we know when it has this ability?

Evaluating Affect Recognition

Here is one test a computer should pass if it can recognize affect. A digital video of a person, containing one or more modalities of expression such as voice, face, gesture, or gait could be observed by both a group of humans and a group of computers. The humans and computers are asked what emotion the person in the video is expressing. When the group of computers and the group of humans respond with the same distribution of answers, then we could say that the computers are recognizing emotions as well as the humans.[1]

Consider if 70% of the humans watching the video think it reveals anger and 30% think it reveals hatred—then a single computer should not be penalized if it recognizes the expression as hatred. When 70% of the computers recognize anger, and the rest recognize hatred, then they have succeeded in matching the humans for this data. Success could be equivalently determined if one computer recognizes that the emotions are "anger with 70% probability and hatred with 30% probability." This test can be repeated for a palette of emotions and subjects.

Alternatively, suppose that we could insure that the person in the video is truly expressing one particular emotion, such as anger. In this case, we might not claim success unless the computer recognized anger. This would be an unusually lofty goal for a computer, however, as most people are not 100% successful at recognizing emotions, even when the emotions are truthfully expressed. In humans, ability to recognize emotions is a sign of emotional intelligence, and can be improved with practice, but perfect performance is never guaranteed.

The test can also be run in two modes—person-dependent, and person-independent. In the person-dependent mode, the computer already knows the person, and customizes its recognition abilities to use what it knows about that individual. Its best recognition performance should occur when it sees the person expressing an emotion that it has seen him express before. In the person-independent mode, the computer may never have seen the person before, and must use some generic recognition abilities. Depending on how closely this new person's expressions are to ones the computer has recognized before, the computer will have correspondingly better success in recognizing his expression and underlying emotion.

I have over-simplified the test scenario, and will postpone discussion of the technical difficulties until I describe how computers are learning to recognize facial and vocal expressions. However, the gist of the test should be clear: for a computer to imitate human recognition ability, we should be able to momentarily swap a computer for a human, and the computer should recognize the same emotions that the human would recognize.

It is entirely possible that computers may become better than some people when it comes to recognizing emotions. This raises the blood pressure of many people, especially those who fear that machines may find out things about them that they wish to keep private. The possibility of computers becoming good at emotion recognition has both positive applications (Chapter 3) and somewhat ominous ones (Chapter 4). But for now, however, numerous advances are needed before computers will be as good as humans at recognizing human expressions of emotion.

Computer recognition of *computer* emotion may also become important. In that case, if computers imitate the same patterns of expression as humans, then recognition can proceed in the same way. If they do not follow human patterns, and they need not do so, then recognition of computer emotion may occur in ways we have yet to conceive. Indeed, it is hard to think of examples of non-human emotions that computers might develop; but, this is a possibility that needs to be allowed if a designer chooses to let a computer evolve its own emotions.

The prospect of computers that recognize our emotions raises another issue: which emotions? Those that are expressed publicly, or those that are expressed via more personal contact? Public emotions are communicated through facial expression, vocal inflection, and overt gestures or body language; these are the more "visible" forms of sentic modulation. They are also the forms over which we have the most control. However, if somebody holds your hand, then they may also feel your pulse racing and sense your clammy hands gradually relaxing and warming. The more personal the contact they have with you, the more likely they are to sense physiological signals from you that are sincere indications of your emotions, signals that most people cannot control by will.[2] Such personal expression is reserved for close friends, lovers, and occasionally physicians or psychiatrists. Because people have frequent physical contact with computers, computers are in a unique position to sense affective signals that are personal, as well as to perceive those that are public.

Differences in Human and Computer Recognition

Computers, not yet of flesh and bone, perceive their world through cameras, microphones, keyboards, mice, and other sensors. These are their eyes, ears, hands, and skin. A computer's sensors may be structured to produce human-like functioning—e.g., juxtaposing two cameras for binocular visual input, shaping pressure sensors onto a robotic hand, or surrounding microphones by an artificial pinna modeled on the human ear. Such efforts have the worthy potential of aiding sensory-impaired humans as well as sensory-impaired machines.

However, machines need not be limited to human-like sensors. Neither need humans be limited, for that matter. A computer could have infrared vision, for example. A human, wearing this computer, can also have infrared vision, given some tools that transform the infrared imagery to visible bands and present it to the human eyes. A computer could have sensitive hearing abilities, outside the normal human range. It could, with some transforming, convert the signals to a form that can be interpreted by a human. The computer might have access to your electrodermal response, pheromones, brainwaves, electromyogram, or blood pressure. It need not be limited to seeing facial and gestural expressions, feeling hand temperature, and hearing vocal inflection, but could pick up signals via any sensors you choose for interacting with it. In this way the computer can have more senses available to it than a person ordinarily has. Consequently, it is possible that computers might recognize emotions and other states that humans would not ordinarily recognize.

In fact, it is possible for computers to recognize affective states that do not presently have names. For example, suppose that late in the afternoon Chris sometimes gets in a peculiar mood where he cannot concentrate on work and finds it relaxing to play 30 minutes of games on the Web. It may be that his physiological signals form a characteristic pattern when he gets in this state. There may be no name for the state he is in. Nonetheless, it may occur with enough regularity that it is useful for the computer to identify its physiological pattern, and to notice that it is a good predictor of some of Chris's behaviors. If typically this state precedes Chris's request for some game software, then the next time the computer detects this state it might preload this software, saving time if Chris decides to play.

Affect Recognition, without other Affective Abilities?

There are numerous advantages to having a machine that can recognize emotions. The computer tutor in the introduction of this book was but one case where it would be useful to recognize the user's emotions and modify the computer's behavior in response. In fact, it may be advantageous to have *only* this affective ability in the machine, i.e. the ability to recognize emotions, without having to "have" emotions, or express them. In fact, the computer might be limited to just recognizing emotional *expressions* such as facial expressions. Chapter 6 describes a system that does this.

However, if its recognition abilities are to pass a test of comparison to human recognition abilities, then other affective abilities are also necessary. People do not recognize emotions based just on the signals seen, heard, or otherwise sensed; they also use higher-level knowledge and reasoning about goals, situations, and preferences. Sarah may enter the stage wearing a smile that might cause most people to recognize her as happy. However, if you know that she just found out that a dear friend committed suicide, then you might label her emotion as deep sadness or grief, and maybe anger, but masked with a smile.

People also reason about the sincerity of emotions, and whether or not a situation calls for sincere expression, or masking of one's true feelings. A certain amount of emotional savvy is helpful for recognizing these situations. This may be in the form of common-sense rules, such as "adults are usually less expressive than children" or in the form of social display rules such as "it is inappropriate to snicker at a funeral." Knowing these rules influences what we think we hear and see—e.g., that snickering sound at the funeral probably was not laughter, but somebody with a head cold trying to breathe. Emotional savvy also needs to be learned, continuously, for individuals and contexts; for example: "Every time Beth has visited Amy and talked about their children, it has cheered Beth up," and "Joe tends to deal with frustration by going off to play solitaire." Such knowledge can be turned around to infer emotional states in the absence of complete information—to guess that after Beth's visit, she will feel better, or that if Joe is playing solitaire, perhaps he was feeling frustrated.

Perhaps the sneakiest influence on human emotion recognition is the mood of the person doing the recognizing. As described earlier, if someone is in a bad mood and sees an ambiguous facial expression, then he or she is more likely to judge the face as negative. In contrast, if the viewer is in a good mood, then the same facial expression is more likely to be seen positively. What gets recognized—a happy or sad face—is determined in part by the mood of the perceiver. Consequently, for computer recognition to imitate human recognition, computers also need to imitate the influence of mood. If the computer has a mood, then its recognition can be biased accordingly. If not, then the computer can be equipped to reason about these biases.

Certain influences of human emotions, such as biasing perception, may seem unnecessary and undesirable to imitate in a computer, except perhaps for the purpose of modeling human emotions for better understanding them. Sometimes the influence of emotions on perception is good, such as to focus the perceiver more closely on a threat, while at other times it is bad, such as when a mood interferes with accurate perception of information. In a human emotion system the good seems often to be inextricably accompanied by the bad. This issue will arise several times in this book: Can we have the good without the bad? There are many ways in which it would be nice to imitate the human emotion system in computers, but there are also ways in which we might want to deviate from it.

Summary of Criteria for Recognition

Design criteria for a computer that can recognize emotion are summarized as follows:

- Input. Receives a variety of input signals, for example: face, voice, hand gestures, posture and gait, respiration, electrodermal response, temperature, electrocardiogram, blood pressure, blood volume pulse, electromyogram, etc.
- Pattern recognition. Performs feature extraction and classification on these signals. For example: analyzes video motion features to discriminate a frown from a smile.
- Reasoning. Predicts underlying emotion based on knowledge about how emotions are generated and expressed. Ultimately, this ability requires perceiving and reasoning about context, situations, personal goals and preferences, social display rules, and other knowledge associated with generating emotions and expressing them.
- Learning. As the computer "gets to know" someone, it learns which of the above factors are most important for that individual, and gets quicker and better at recognizing his or her emotions.
- Bias. The emotional state of the computer, if it has emotions, influences its recognition of ambiguous emotions.
- Output. The computer names or describes the recognized expressions, and the emotions likely to be present.

Embedded in these criteria are numerous technical requirements, for example, "receiving inputs" requires accurate technology for gathering digital physiological, audio, and visual signals, as well as research to determine which signals are most important for the task at hand. In pattern recognition, informative features of the signals need to be identified—statistical, structural, nonlinear, etc.—together with conditioning variables that influence the meanings of these features. Part II will delve more deeply into the implementation of emotion recognition.

Computers that Express Emotions

The human voice is always changing. Even if the receptionist says the same "Good afternoon, welcome to Sirius Cybernetics Corporation," every time you call, his or her intonation is always slightly different—a cheerful hello, a brusque hello. The same person can speak the same words, but say them entirely differently. Sometimes the part of the message that communicates emotion is the most important part.

One problem the information age has brought is that of too much information, which tends to lead to cognitive fatigue and a reduced ability to accurately process new inputs. In contrast, information presented through the "affective channel" does not usually demand conscious attention. Affective information can be perceived in parallel with non-affective information, without increasing your workload. In speech, *what* is said can be considered the semantic information, and *how* it is said can be considered largely the affective information. The latter is communicated through the modulation of vocal parameters in ways I will describe more carefully later. Speaking a simple "Hello!" in a happier tone than usual is, for both the speaker and listener, less work than speaking the two separate messages: "Hello" and "I'm more happy than usual at this moment."

If the computer spoke an audible greeting, its voice would become a channel that could be used to express important information about valence or urgency. Its greeting might ring of joy or sadness—perhaps a pre-indicator of the affective tone of the news that it found for you while you were asleep last night. In this example the communication accomplishes two things at once: greeting you, and informing you about the nature of news, without demanding extra time on your part. Affective inflection not only makes for a more pleasant interaction, but also it makes for efficient communication.

A computer can express emotion without having emotion, just like humans can express emotions that they do not have. A software agent acting as a tour guide could post a happy greeting without any underlying emotional state. A computer switchboard operator could vary the affective quality of its greeting in an effort to sound more pleasing. It may seem peculiar that one could have emotional expression without having emotions to express. Nonetheless, this is entirely possible in a computer. The basic requirement for a computer to have the ability to express emotions is that the machine have channels of communication such as voice or image, and an ability to communicate affective information over those channels. For example, a computer or software agent that displays a face could use an expression such as a smile—like a Macintosh does, upon booting up. Alternatively, if a computer actually had emotions, then it might also have some direct readout display of its emotional state that changed, like a human smile, every time it encountered a suitable stimulus, like a funny joke. In these simple ways, computers already have channels for expressing emotions, whether or not they have emotions.

A machine can be used as a channel for transmitting human emotions. When two humans communicate via email or via teleconference, the machine and network act as a communication channel connecting people. Typically the channel is band-limited: all the information it receives at one end cannot be sent to the other end; some of it is lost. For example, it might convert a speech signal to text, throwing away the affective part of the signal. We might describe the *affective bandwidth* of a channel as how much affective information the channel lets through. *Affective information* might refer to the entropy of the affective part of the signal, which indicates how many bits are required to describe the part of the signal that carries the affective message. The idea of entropy as information is that of Claude Shannon, whose renowned theory of information applies also as a framework for emotion communication (Shannon and Weaver, 1963; Buck, 1984). Using the framework of information theory, we might also speak of the relative "affective channel capacity" of various forms of communication, to indicate how much affective information can be carried.

When sending the same words, different channels allow for more or for less affective channel capacity: email usually communicates the least affect, phone slightly more, videoteleconferencing more still, and "in person" communication the most. It is usually assumed that technology-mediated communication always has less affective bandwidth than person-to-person communication. Sometimes the limits on affective bandwidth are desirable. You might wish to choose a medium where your emotions are not as easily seen. However, rarely is it desirable to have these limits forced on you. Affective recognition and expression can be used to allow for more possibilities in communication, even with limited technology. For example, if there is not enough bandwidth to transmit each person's facial expression, the computer might recognize the expression, send just a few bits describing it, and then represent these with an animated face on the other side of the channel.[3]

Might technology increase affective bandwidth? Virtual environments and computer-mediated communication offer possibilities that we do not ordinarily have in person-to-person communication. Potentially, communication through virtual environments could provide new channels for affect—perhaps, as one idea, via sensors that detect physiological information and relay its significant information. In this way, computer-mediated communication might potentially have *higher* affective bandwidth than traditional "in person" communication. Of course, the use of such communication would not be desirable all the time. It might be saved for long-distance communication with loved ones, or for contact with emergency medical personnel.

Consider if a computer could perceive its user typing happily while sending email. If desired by the sender, who ultimately should have control over what gets sent, the computer could pass some of the affective quality of the interaction through the channel to the recipient. But, how would it express this on the other side? Would it transmit the sound of the keys clicking? Or, use a bright sounding voice to convert the email text to speech? Would it change the color of the email window to reflect a sunny mood? If the email sender were furious, and wanted the recipient to know, then even more possibilities arise, since the channels of emotional expression available to a machine are different than those available to a person. The machine might literally vent its heat in the direction of the recipient.

Evaluating Affect Expression

One way to test a computer's ability to express emotions is to have it express certain emotions and have a human try to recognize what it has expressed. Clark Elliott, of DePaul University, conducted one instance of this test. The computer used facial and vocal inflection to communicate several emotional states. Elliott videotaped the computer's efforts, and also videotaped an actor saying the same sentences trying to communicate the same emotional states. The sentences were deliberately ambiguous, so as not to reveal the emotions by their semantic content. In tests with human listeners, the humans were more accurate at recognizing the emotions when expressed by the computer than when expressed by the actor. The computer succeeded roughly 70% of the time in conveying the intended expression, while the actor succeeded only about 50% of the time. The conclusion is that not only can computers express emotions, but in some cases, they can do so more accurately than humans, especially when they employ caricature (Elliott, 1997).

Another possibility for evaluating emotional expression by a computer is to test for the computer's ability to *induce* emotion, by contagion. Goleman, in his book *Emotional Intelligence*, gives an example of a steamy August New York City day when he dragged his sullen self onto a bus, only to encounter a cheery bus driver, who greeted and entertained his passengers with such enthusiasm and genuine joy that their moods were nearly all transformed by the end of their ride. We all know cases where another person's emotional expression has altered our own. Computers also have an opportunity to influence the mood of their users, negatively or positively. It is possible that a computer, which expressed an underlying mood, might contagiously transmit its mood to people.

Differences in Human and Computer Expression

Not only do computers have the ability to physically vent heat, but computer expression of emotions can differ from that of humans in that there may be different emotions to express. Computers may have unique affective states, with correspondingly unique forms of expression, perhaps decodable only by other computers. These might be meaningless to most of us: "Warning: emotion 90. Free-memory cycle stress," but may nonetheless communicate important information to another computer, such as a lack of receptivity to large new memory demands at the moment.

A computer will need to know when to be subtle in its emotional expression, and when to exaggerate to make an emotion clear. Actors and animators know well the importance of being unambiguous in communicating emotion. The rest of us are often not as good at this. Computers endowed with such abilities might be more effective than their users at expressing emotions. In fact, in some applications, such talented computers might be used as amplifiers or coaches, helping a person to more accurately communicate emotions.

A problem with machines that can express emotion, but not have emotion, is that this case does not occur with humans. Humans have emotions whether or not they express them. Computers do not have to have emotions to express them—they can just paint their screen with a smile, if that's what the program says to do. A computer could be in one internal affective state, while expressing a different state. Somebody who looks at an agent's face might presume the agent was expressing its affective state. Even if the computer had separate systems for displaying intentional and spontaneous emotions, these need not be implemented the same way they are in humans. More than two such systems might exist, and so forth. In fact, the whole question of "intentional" vs. "spontaneous" expressions in a computer is an interesting one, for an observer may never be able to insure that a spontaneous-appearing expression was not artificially created. Chapter 4 will look at issues of computer deception, together with other potential problems that need to be addressed with the design of affective computing systems.

Affect Expression, without other Affective Abilities?

A computer can have the ability to express emotions without "having" them or recognizing them, as I mentioned above with the examples where the computer simply acts as an unemotional channel for transmission of human emotion, or where a computer generates expressions out of social politeness or to efficiently convey the affective tone of information. Humans can also express an unfelt emotion, such as making an intentional smile vs. smiling because something is funny. In humans, however, the bodily expression feeds back and can cause an emotion to actually be felt, as we read about earlier in experiments where people's faces were posed into an expression, causing them to feel the corresponding emotion. The ability in humans for emotions to co-occur with their expression makes it especially difficult to separate "expressing an emotion" from "having an emotion." In particular, a human who is strongly experiencing one emotion can find it hard to express a different emotion. A person caught up in an angry rage, when asked to express love, will not succeed in a very convincing expression of love until the anger has subsided. Having some kinds of emotions makes others hard to express. This bias-exclusion effect impacts a system that can both express and have emotions.

Summary of Criteria for Expression

Design criteria for a computer that can express emotion are summarized as follows:

- Input. Computer receives instructions from a person, a machine, or from its own emotion-generation mechanisms if it has them, telling it what emotion(s) to express.
- Intentional vs. spontaneous pathways. The system may have at least two paths for activation of emotional expression: one that is intentional, and one that is spontaneous. The former is triggered by a deliberate decision, while the latter acts within a system that has emotion, automatically modulating some of the system's outputs with the current emotion.
- Feedback. Not only does affective state influence affective expression, but the expression can influence the state.
- Bias-exclusion. It is easiest to express the present affective state, and this state can make the expression of certain other states more difficult.
- Social display rules. When, where, and how one expresses emotions is determined in part by the relevant social norms.
- Output. System can modulate visible or vocal signals such as a synthetic voice, animated face, posture and gait of an animated creature, music, and background colors, in both overt ways such as changing a facial expression, and in subtle ways such as modifying discourse timing parameters.

Computers that "Have" Emotions

Can machines feel? This question may be the most profound one within the topic of affective computing. Feelings are often considered to be that which separates human from machine. Furthermore, the possibility of creating a computer with subjective feelings hinges on the issue of machines having consciousness, which is itself a topic of unfettered debate. What does it mean, computationally, for a computer to "have" emotions? How would we know if a computer did or did not? Despite the tremendous complexity of these questions, I will propose an answer below by describing five components of a system that has emotions. The emphasis on five components is for explanatory purposes only and is not intended to imply how the mechanisms of emotion are structured in an affective system. In a healthy human emotional system, all of the five components are present. Although all the components are present in a complete emotional system, they do not have to be all activated at every moment the system is operating. I suggest that we may argue that a computer "has emotions" when all five components are present in a computer.

Component 1: Emergent Emotions and Emotional Behavior

The first component of having emotions is what I will call "emergent emotions." Emergent emotions are those which are attributed to systems based on their observable emotional behavior—especially when the system which is behaving has no explicit internal mechanism or representation for emotions. In humans, who *do* have internal emotional mechanisms, this component refers to emotional behaviors and other outward expressions.

An example of emergent emotions comes from Braitenberg's *Vehicles* (Braitenberg, 1984). One of his simplest vehicles has two light sensors and two motors, hooked up so that when it sees a light source straight ahead, it moves toward it, and bangs into it, hitting it frontally. When the source is not straight ahead, then the vehicle turns and moves so that it still approaches the source and hits it. Its behavior is seen by human observers as aggressive, as if the vehicle felt a negative emotion toward the light source. Vehicles can also be wired so as to linger near the source and not damage it. This behavior gives the impression of a more favorable emotion, such as love.

In 1962, Masanao Toda, a Japanese psychologist who emphasizes the importance of studying whole systems, including perception, action, memory, and learning, proposed a scenario with a "fungus eater," a humanoid robot, to illustrate how emotions would emerge in a system with limited resources operating in a complex and unpredictable environment (Toda, 1962). Toda's robot has a goal of collecting as much uranium ore as possible, while regulating its energy supply for survival. The robot has rudimentary perceptual, planning, and decision-making abilities. With the inclusion of "urges," which Toda defines as "motivational subroutines linking cognition to action," Toda argues that the robot would become emotional. According to Toda, urges come in two flavors: emergency urges, such as fear, and social urges, such as love. Urges are triggered in relevant situations, and subsequently influence cognitive processes, attention, and bodily arousal. Toda's proposal has influenced the AI community by emphasizing the emergence of emotions in complex goal-directed autonomous systems.[4]

Simply expressing an emotion can be seen as a kind of emotional behavior. In this way a machine that expresses emotions, even without having them, might have emotions attributed to it. The Macintosh computer is a simple example: when the user boots up the system, its little disk face smiles. The user might rationalize that the Macintosh has the goal of interacting with her and is therefore "happy" because its goal is achieved. In this sense it might be said to have emergent emotions since it gives the appearance of having emotion without having any explicit internal mechanisms that produce emotion. Nonetheless, like Braitenberg's vehicles, the Mac has no internal emotions.

Component 2: Fast Primary Emotions

Humans can be startled, angry, or afraid, before the signals even get to the cortex, and before becoming aware of what is happening. In fact, sometimes we are already behaving, such as jumping out of the way of danger, before we become aware of an emotion, such as fear. Many animals have these essentially hard-wired, innate responses, especially to potentially harmful events. Such emotional responses may not always be accurate; for example, you might also jump out of the way of a non-dangerous stimulus. Fast emotions tend to err on the side of survival.[5]

These innate, quick and dirty reactions are what Damasio calls "primary" emotions, and probably include at least fear, surprise, and anger. As each emotion is studied and its circuitry is deciphered, the criteria for primary emotions are likely to become more specific, and the list more complete. In the meantime, the findings of neurologists are revealing how these emotions are implemented, and therefore, how each might be imitated in computers.

In particular, for the fear emotion, the fundamental mechanism seems to be a danger detection system—one that operates primarily in a loose but quick "jump out of the way" sense, and only secondarily in a more precise but slower evaluative sense: "Oh, that was nothing to be afraid of" or "Whew! I could have been crushed!" Fear can also arise cognitively, perhaps while ruminating about a future threatening event, but that can be considered a slower cognitive process.

The brain mechanisms for primary emotions point to two communicating systems—a rough pattern recognition system that acts fast, and can "hijack" the cortex, but often makes mistakes by triggering false alarms, and a finer pattern recognition system that is slower but more precise. Together, they work to detect important events, such as danger, and to trigger the subjective feelings, sentic modulation, regulatory controls, and behaviors that are crucial to survival. In particular, primary emotions appear to be important for re-allocation of limited resources, as illustrated in this robot scenario:[6]

A robot exploring a new planet might be given some basic emotions to improve its chances of survival. In its default state, it might peruse the planet, gathering data, analyzing it, and communicating its results back to Earth. However, suppose at one point the robot senses that it is being physically damaged. At this instant, it would change to a new internal state, perhaps named "fear," which causes it to behave differently. In this state it might quickly re-allocate its resources to drive its perceptual sensors. Its "eyes" might open wider. It would also allocate extra power to drive its motor system so that it could move faster than usual, to flee the source of the danger. As long as the robot remains in a state of fear, it would have insufficient resources to perform its data analysis, like humans who are unable to concentrate on other things until the danger has passed. Its communication priorities would cease to be scientific, changing to a call for help, if it knows help might be available. The "fear" state would probably remain until the threat passed, and then its intensity would decay gradually, until the robot returned to its default state where it could once again concentrate on its scientific goals.

When thinking about implementing primary emotions in computers, it is tempting to wire them so that they always trigger a certain behavior in response to certain stimuli. However, this is not necessary, and is probably not the way they are implemented in humans and other animals. For example, for the primary emotion of fear, it is true that a mouse might fear a cat and flee in its presence. However, if the only food source is near the cat and the mouse gets sufficiently hungry, then the mouse may approach the food despite its fear of the cat. Although the fear response that primes the mouse to flee may be innate, the resulting behavior can be influenced by motivational and other factors.

Component 3: Cognitively Generated Emotions

A third component of having emotions is what I will call cognitive emotions, to emphasize the explicit reasoning typically involved in their generation. Cognitive-appraisal emotion theorists have tried to explain the generation of large sets of emotions based on the answers to questions such as, "Was the situation giving rise to the emotion certain or uncertain?" "other-caused or self-caused?" and so forth. Some have even surmised that all emotions are cognitively generated, even emotions that we now know are generated in the limbic system before the signals from their stimuli arrive in the cortex. Upon learning about the latter, some theorists have emphasized that cognitive appraisal can happen not just consciously, but subconsciously, perhaps even by structures in the limbic system. However, I think it is important to maintain the distinction between fast primary and slower secondary emotions, keeping the former in Component 2. Therefore, this third component includes only those emotions believed to be initiated by cortical reasoning, such as when we assess that someone has impeded our goals, and subsequently feel anger toward them. In contrast with the primary emotions, cognitively generated emotions may arise slowly, possibly as a consequence only of one's deliberate thoughts. If merely thinking, "Don't worry; be happy" causes you to be happy, then that happiness is cognitively generated.

In a healthy human, cognitively generated emotions usually provoke an emotional experience with subjective feelings, especially if the emotion is intense. Damasio talks about "secondary emotions" as the ones that are cognitively generated but subsequently activate limbic responses and bodily feelings. However, in a computer, an affective state need not be accompanied by feelings or physiological components of emotion. We see a similar situation in Damasio's patients who can reason about emotions cognitively without any of the normal accompanying feelings. When Elliot saw an image of a bloody mutilated face, he knew—cognitively—that it was horrific. But it was a cool rational horror. Elliot knew that before his brain tumor, he would have also *felt* terrible seeing the mutilation image. His cognitive generation of emotions was fine, including knowing what kind of feeling they should generate. Computers, like Elliot, can generate cognitive emotions without their normal accompanying feelings.

Ortony, Clore, and Collins wrote in their influential book, *The Cognitive Structure of Emotions*, that it is not important for machines to have emotions, but it is important to AI that computers be able to reason about emotions—especially for natural language understanding, cooperative problem solving, and planning. I will underscore some of their remarks: certainly it is important for machines to be able to reason about emotions. To the extent that computers interact with people, such reasoning abilities are required for them to interact more sensitively and intelligently. However, we know now that computers with only the ability to reason about emotions, cognitively, are similar to the emotion-impaired patients. Cognitive reasoning is important for intelligent decision making and behavior, but it is not always sufficient.

Cognitive reasoning is the way emotions are most frequently generated in machines today, especially in animated software agents. Typically, a set of rules is constructed to generate emotional states given certain inputs. Ironically, the most frequent paradigm for generating these emotions is based on the Ortony, Clore, and Collins (OCC) structure, despite the fact that they did not have emotion generation in mind when they formulated their cognitive appraisal theory (see Chapter 7). Via cognitive reasoning, the computer can deduce that a sequence of events causes an emotion to arise. By the same reasoning, applied to its personal events, it can cause an emotion to arise within itself. Hence, we see that the OCC structure can be used not just for reasoning about emotions, but also for generating them.

Component 4: Emotional Experience

Consider if one of Braitenberg's vehicles were given the ability to recognize and label its retreating behavior as "fear." Moreover, suppose that this ability was extended so that it could recognize and label all its emotional behaviors. The vehicle could then be said to have a rudimentary awareness of its emotional behaviors. If the vehicles had an internal emotional mechanism that they could similarly become aware of, then they might be said to have an awareness of their internal emotions. The ability for a system to be cognitively aware of its emotional state is a key aspect of emotional experience, although the system need not be aware of its emotions at all times.

There is a second aspect of emotional experience, and that is the awareness of its physiological accompaniments. For a human these include heart beat, increased perspiration, a readiness in one's legs to run, trembling, cold feet and so forth. For the vehicles, physiological factors might correspond to sensing its motors' speeds and its motion characteristics—turning, shaking, and so forth. Most computers do not have sensors that could discern their physical state at a given moment, but these could be added. For example, in one of our experiments, we have given a computer the ability to sense the moment when its display receives an image, by measuring voltage changes on the wires hooked to its monitor. Nevertheless, since computer physiology differs from human physiology, with different senses, a computer's physiological experiences of emotion will differ from those of a human.

There is a third, and final aspect of emotional experience, which is perhaps the trickiest to understand, and consequently the hardest to implement in a machine. Moreover, it is the most familiar aspect of emotion. This is the internal *subjective feeling* or "gut feeling" that leads you to know something is good or bad, that you like it or dislike it. It is unclear precisely what constitutes these feelings. People often think of them as visceral, and indeed, there are hormones released by the viscera that travel through the blood to the brain. But scientists know that the body organs which release these hormones consist of "smooth muscle" that responds much more slowly than the striated muscles in the somatic system, such as those in facial expressions or skeletal movements. Their slow speed makes it unlikely that the viscera are giving rise to an emotional response, although they may be contributing to the overall emotional feeling. On the other hand, neurotransmitters act rapidly in the brain, and their activation likely initiates these subjective feelings. When biochemical substances become easy to measure and observe during emotional arousal, then a physiological explanation may be found for this aspect of emotional experience, so that these last two aspects of emotional experience—physiological awareness and subjective feelings—can be combined into one. At this point, scientists might even describe this aspect as "objective feelings," meaning that the feelings have observable physical components, which can reliably predict the labels that a person assigns to the feelings; however, this remains to be shown empirically.

In summary, emotional experience consists of:

- Cognitive awareness
- Physiological awareness
- Subjective feelings

This triptych may seem to be a needless aspect of emotion to give to computers. It is possible to give computers emotional behaviors, fast primary emotions, and cognitively generated emotions without giving them an emotional experience, so why bother giving them this component? This question is a close cousin of another: "Why do humans have emotional experiences?" In searching for its answer, we run into the issue of consciousness. I deal with this in its own section later, but let me offer one brief hypothesis here. My hypothesis is that an emotional experience gives us the ability to better understand, learn about, and regulate our activities—for example, why we want certain things, what has and has not met our needs, and how we might do things better. Knowing you are in love can explain much silliness, whether or not it justifies it. Noticing how tense your muscles are can inspire an effort to reduce stress or anger. Feeling that something is wrong can provoke you to re-examine a situation. Experiencing emotions not only adds a special quality to life, but also helps explain much of our mental and physical behavior, for example why we linger on some thoughts, e.g., it feels good to think about what she said, and why we introspect and make changes regarding others, e.g., it feels bad not to have returned her call so you decide to apologize. Through emotional experience we gain insight into our own motivations and values; we become able to better understand and utilize the powerful influences emotions exert.

Component 5: Body-Mind Interactions

We have seen in the previous chapter that emotions influence decision making, perception, interest, learning, priorities, creativity, and more. Emotions influence cognition, and therefore intelligence, especially when it involves social decision making and interaction. Furthermore, not only do human emotions influence cognition, but they influence other physiological systems besides the brain—modulating vocal and facial expressions, influencing posture and movement, even influencing digestive processes and the workings of the immune system. What you think, what you eat, what medications you take, your posture, and more can influence your emotions. Emotions intricately interact with the human body and mind.

Scientists know that emotions influence memory and memory retrieval, and it may be that this influence on memory is at the root of all of emotion's influences on cognition. In fact, it may later be found that the internal *feelings* discussed above in Component 4 are the causal agent behind these influences, and that therefore these influences would not happen in a system that had no subjective feelings or facsimile thereof. As mentioned earlier, all five of the components discussed here are present in humans, and it is unclear precisely how they interact. For now, we do not know the mechanism that gives rise to the influences that I lump into this fifth component.

Not only does emotion influence cognitive and bodily functions, but emotion is, itself, influenced by them. I emphasized above that cognitive thoughts, which include concerns, goals, and motivations, can generate emotions. Similarly, biochemical processes such as changes in hormones and neurotransmitter levels, physical drives such as hunger and low blood sugar levels, and physical feedback, such as facial posing, can influence emotions and their generation. Prozac and a number of other depression medications adjust the uptake of serotonin, a neurotransmitter, in an effort to alleviate a depressed mood. Furthermore, the precise ways in which biochemical processes interact with emotion may differ from person to person; personality, innate differences in temperament, and learned differences in emotional development all play a role in these influences.

The aspect of emotion's interaction with the mind and body that may be the most important for computers is the influence of emotion on cognitive processes—especially the aspect that is missing in Damasio's patients, which is crucial for intelligent decision-making in many situations. In humans, emotion also influences flexibility, creativity, and learning. Consider the following scenario of a smart personal assistant:[7]

A computer software agent is learning to be a smart personal assistant, to aid you in scheduling meetings and retrieving important information. It has two ways of getting feedback—direct or indirect. In the direct case, you select preferences on a menu, effectively programming it. In the indirect case, it watches how you respond to it and adapts itself. Suppose that it enters a state called "feel good" when (1) you express pleasure at its performance, and (2) when you succeed at a task more efficiently and accurately than usual. Additionally, it might have a corresponding "feel bad" state for the reverse of these, as well as a neutral "no emotion" state, a "feeling curious" state, and an "I'm puzzled" state. These states influence a variety of things, such as the values of items added to memory, and their associations with things good and bad, important or minor. When the system has been in a state of "feel good" for several days, it might become more curious and decide to try out new ideas for helping you, taking more risks. If it lingers in a "feel bad" state, it might allocate more resources to introspection, to trying to figure out if it could do something differently and more effectively. If it interrupts you and you look upset for being interrupted, it would trigger a small bad feeling, marking this situation as aversive, biasing it away from repeating this situation and prompting it to try to understand why you did not like being interrupted at this particular time. When presented with a complicated set of demands, it will weigh the "feel good" and "feel bad" associations that exist and try to choose an action that helps satisfy goals (1) and (2). It will aim for a dynamic balance—recognizing that often you will not show pleasure even though it is performing well, and that sometimes you will complain or show approval inconsistently, regardless of what it does. At such times, depending on how calm or agitated you are, as measured from your norm, it can either ask you for clarification, or make a note to come back later and try to understand this situation, perhaps when you are not so agitated. When it is overloaded with so much information that it cannot consider all the possibilities in time to give you the decision you need, then it uses the valenced markers it has learned to guide it toward incorporating the most significant information in the decision. Its use of emotions helps it manage information overload, regulate its prioritization of activities, and make decisions more flexibly, creatively, and intelligently.

The imitation in a computer of emotion's influence on the human body is less obvious. Although it might be reasonable for a computer's emotions to automatically modulate its available modes of expression, imitating human sentic modulation, computers have no present obvious need for imitating biochemical or electrochemical influences of emotion. Furthermore, no computers have immune systems that function like human immune systems, even though many of them are equipped with mechanisms that watch for certain kinds of computer viruses, and attempt to deal with those viruses. However, as computer systems become more complex and take on greater responsibilities with more real-time constraints, we may find that their regulatory mechanisms will function more like the physical regulatory mechanisms that interact with a human emotional system.

Before concluding this section, let me restate that it is important to keep in mind that all computers will not need all components of all emotions. Just like simple animal forms do not need more than a few primary emotions, all computers will not need all emotional abilities, and some will not need any emotional abilities. Humans are not distinguished from animals just by a higher ability to reason, but also by greater affective abilities. We can expect more sophisticated computers to need more sophisticated emotional abilities.

My claim, which opened this section, was that all five components occur in a healthy human, and if a computer has all five components of emotions, then it can be said to have emotions. However, let me add that not having all five does not imply that the system does not have emotions. For example, there are humans who are so handicapped that although they have emotions with respect to Components 2–5, they may not be able to communicate them with their behavior, as per the first component. We would not say that these people do not have emotions, but rather, that it is uncertain what they are experiencing since they cannot communicate. Another example is

Damasio's patients, who lack part of emotional experience (Component 4), and some of the influences of emotion on decision-making (Component 5), but still have emotional behaviors, primary emotions, and the ability to reason about emotions cognitively. They have an *impaired* emotion system; it is not accurate to say that they have no emotions.

Finally, some might argue that all the components above specify functions of emotions, and that there are other ways to fulfill these functions than by giving emotions to computers. I have been asked several times, "Can't we accomplish all of these desirable abilities without having to give the machine emotions?"

The answer is "yes" in theory, for we know that any mathematical function can be constructed in more than one way; therefore, why not any kind of function, including any human function? In practice, however, suppose that we tried to build into a computer all the functions and influences of emotion, one by one, without giving it emotions *per se*. Each influence of emotion would be implemented by some mechanism that gives rise to that influence. For example, there would be a mechanism to give rise to influences on learning, another mechanism to give rise to influences on decision making, another to regulate the use of physical resources, another to give rise to influences of motivation, another to provide a fast response in the presence of a threat, and so forth, implementing the many beneficial influences of emotion without implementing emotions. In so doing, we would no doubt notice that some of these mechanisms might be combined in clever ways, to streamline the system and make it robust. By the time we have insured that all these influences are in place in a sufficiently general, flexible, and efficient manner, however, then will we not have created a system of mechanisms that is effectively an emotion system? In other words, we may find that building in the positive benefits of emotions is not all that different from building in emotions. Of course, it does not have to be called "emotions," especially since that word comes loaded with many negative connotations. Nevertheless, that is essentially what the result would be.

Evaluating Performance

The five components of emotion give rise to separate tests to evaluate if the computer has achieved each of them. In particular, we can adapt human tests, substituting a computer, software agent, or other digital system. Some possibilities include:

Component 1: Emotional behavior. Does an outsider, observing the behavior of the system, describe its behavior with the correct emotional adjective?

Component 2: Fast primary emotions. Does the system respond quickly and with distinctive behaviors to stimuli of a threatening or otherwise urgent nature, mobilizing its resources for suitable responsive actions?

Component 3: Cognitive emotions. Given a situation, can the system reason about the emotions it is likely to generate, and apply the same reasoning to situations it is in, labeling its own state with the generated emotion?

Component 4: Emotional experience. What is the report of the system's internal affective state? Can its sensors discriminate different feelings for different emotions?

Component 5: Body-mind interaction. Is the system more likely to retrieve, learn, and recognize positive information when in a positive mood, and vice-versa for a negative mood?

Additional tests. By taking a human-human situation, and substituting a computer so that it is a human-computer situation, there are many possible tests. For example, if human A is put in a room with a gregarious and ebullient human B, then human A is likely to acquire a more positive mood. If either human is replaced with an affective computer, will the results of the mood exchange be the same? This "contagion test" is just one of many possibilities. Furthermore, this test involves not only generating emotions, but also recognizing or expressing them. Following the paradigm of adapting results from human emotion theory, we can propose a battery of tests for computer emotions.

Summary of Components of an Emotion System

To summarize, a computer can be said to have emotion if it has the following five components that are present in healthy human emotional systems:

1. System has behavior that appears to arise from emotions.

2. System has fast "primary" emotional responses to certain inputs.

3. System can cognitively generate emotions, by reasoning about situations, especially as they concern its goals, standards, preferences, and expectations.

4. System can have an emotional experience, specifically:
   - Cognitive awareness
   - Physiological awareness
   - Subjective feelings

5. The system's emotions interact with other processes that imitate human cognitive and physical functions, for example:
   - memory
   - perception

- decision making
- learning
- concerns, goals, motivations
- attention, interest[8]
- prioritizing
- planning
- sentic modulation
- immune system functions
- regulatory mechanisms

Computers with Minds and Bodies

I am sometimes asked, "Do computers have to have bodies to have emotions?" After all, in humans we know that emotions have both physical and cognitive components. We know of no brains without bodies, except perhaps those in fanciful stories, such as Daniel Dennett's delightful "Where am I?" (1978). Indeed, recent books such as Damasio's *Descartes' Error* refer to the mind-body separation as misguided, and emotions as an example par excellence of mind-body interaction.

It is tempting to say for a computer that the hardware is the body and the software is the mind. Furthermore, a hardware implementation used to be the only way to implement fast processing in a computer, because software was too slow; consequently, one might expect that fast primary emotions would best be implemented in hardware, while slower cognitively-generated emotions might best be implemented in software. However, this speed distinction is no longer very important because of advances in technology. A more important distinction than hardware vs. software, is "fast but rigid" vs. "slow but flexible." We do not find a clear software vs. hardware distinction among the human brain's electrical and chemical reactions. What we find is many separate processes, including some (limbic) that handle emotions in a fast, but rigid way which is prone to make mistakes, and some (cortical) that handle emotions in a slower way, which is more accurate and adaptable.

The distinction between hardware and software is especially blurred when an emotion system is given to a computer that is entirely simulated in software, or to a software agent or animated creature that exists in software with a software body. In these cases, software mechanisms can be used for implementing both quick and dirty emotional processing, and slower more accurate emotional processing. The precedence of the former over the latter can be maintained in software so that the agent would still be capable of an "emotional hijacking" whereby the quicker processes, perhaps with interrupt capabilities, assume control and the agent responds rapidly, albeit not always accurately. For example, an animated agent might cease eating to jump out