Emotional Machines

Dave, stop... Stop, will you... Stop, Dave... Will you stop, Dave... Stop, Dave... I’m afraid. I’m afraid... I’m afraid, Dave... Dave... My mind is going... I can feel it... I can feel it... My mind is going... There is no question about it... I can feel it... I can feel it... I’m afraid.

—HAL, the all-powerful computer, in the movie 2001

HAL IS CORRECT TO BE AFRAID: Dave is about to shut him down by dismantling his parts. Of course, Dave is afraid, too: HAL has killed all the other crew of the spacecraft and made an unsuccessful attempt on Dave’s life.

But why and how does HAL have fear? Is it real fear? I suspect not. HAL correctly diagnoses Dave’s intent: Dave wants to kill him. So fear—being afraid—is a logical response to the situation. But human emotions have more than a logical, rational component; they are tightly coupled to behavior and feelings. Were HAL a human, he would fight hard to prevent his death, slam some doors, do something to sur-
survive. He could threaten, “Kill me and you will die, too, as soon as the air in your backpack runs out.” But HAL doesn’t do any of this; he simply states, as a fact, “I’m afraid.” HAL has an intellectual knowledge of what it means to be afraid, but it isn’t coupled to feelings or to action: it isn’t real emotion.

But why would HAL need real emotions to function? Our machines today don’t need emotions. Yes, they have a reasonable amount of intelligence. But emotions? Nope. But future machines will need emotions for the same reasons people do: The human emotional system plays an essential role in survival, social interaction and cooperation, and learning. Machines will need a form of emotion—machine emotion—when they face the same conditions, when they must operate continuously without any assistance from people in the complex, ever-changing world where new situations continually arise. As machines become more and more capable, taking on many of our roles, designers face the complex task of deciding just how they shall be constructed, just how they will interact with one another and with people. Thus, for the same reason that animals and people have emotions, I believe that machines will also need them. They won’t be human emotions, mind you, but rather emotions that fit the needs of the machines themselves.

Robots already exist. Most are fairly simple automated arms and tools in factories, but they are increasing in power and capabilities, branching out to a much wider array of activities and places. Some do useful jobs, as do the lawn-mowing and vacuum-cleaning robots that already exist. Some, such as the surrogate pets, are playful. Some simple robots are being used for dangerous jobs, such as fire fighting, search-and-rescue missions, or for military purposes. Some robots even deliver mail, dispense medicine, and take on other relatively simple tasks. As robots become more advanced, they will need only the simplest of emotions, starting with such practical ones as visceral-like fear of heights or concern about bumping into things. Robot pets will have playful, engaging personalities. With time, as these robots gain in capability, they will come to possess full-fledged emotions: fear and anxiety when in dangerous situations, pleasure when accomplishing a desired goal, pride in the quality of their work, and subservience and obedience to their owners. Because many of these robots will work in the home environment, interacting with people and other household robots, they will need to display their emotions, to have something analogous to facial expressions and body language.

Facial expressions and body language are part of the “system image” of a robot, allowing the people with whom it interacts to have a better conceptual model of its operation. When we interact with other people, their facial expressions and body language let us know if they understand us, if they are puzzled, and if they are in agreement. We can tell when people are having difficulty by their expressions. The same sort of nonverbal feedback will be invaluable when we interact with robots: Do the robots understand their instructions?
When are they working hard at the task? When are they being successful? When are they having difficulties? Emotional expressions will let us know their motivations and desires, their accomplishments and frustrations, and thus will increase our satisfaction and understanding of the robots: we will be able to tell what they are capable of and what they aren’t.

Finding the right mix of emotions and intelligence is not easy. The two robots from the Star Wars films, R2D2 and C3PO, act like machines we might enjoy having around the house. I suspect that part of their charm is in the way they display their limitations. C3PO is a clumsy, well-meaning oaf, pretty incompetent at all tasks except the one for which he is a specialist: translating languages and machine communication. R2D2 is designed for interacting with other machines and has limited physical capabilities. It has to rely upon C3PO to talk with people.

R2D2 and C3PO show their emotions well, letting the screen characters—and the movie audience—understand, empathize with, and, at times, get annoyed with them. C3PO has a humanlike form, so he can show facial expressions and body motions: he does a lot of hand wringing and body swaying. R2D2 is more limited, but nonetheless very expressive, showing how able we are to impute emotions when all we can see is a head shaking, the body moving back and forth, and some cute but unintelligible sounds. Through the skills of the moviemakers, the conceptual models underlying R2D2 and C3PO are quite visible. Thus, people always have pretty accurate understanding of their strengths and weaknesses, which make them enjoyable and effective.

Movie robots haven’t always fared well. Notice what happened to two movie robots: HAL, of the movie 2001 and David, of the movie AI (Artificial Intelligence). HAL is afraid, as the opening quotation of this chapter illustrates, and properly so. He is being taken apart—basically, being murdered.

David is a robot built to be a surrogate child, taking the place of a real child in a household. David is sophisticated, but a little too perfect. According to the story, David is the very first robot to have “unconditional love.” But this is not a true love. Perhaps because it is “unconditional,” it seems artificial, overly strong, and unaccompanied by the normal human array of emotional states. Normal children may love their parents, but they also go through stages of dislike, anger, envy, disgust, and just plain indifference toward them. David does not exhibit any of these feelings. David’s pure love means a happy devoted child, following his mother’s footsteps, quite literally, every second of the day. This behavior is so irritating that he is finally abandoned by his foster mother, left in the wilderness, and told not to come back.

The role of emotion in advanced intelligence is a standard theme of science fiction. Thus, two of the characters from the Star Trek television shows and films wrestle with the role of emotion and intelligence. The first, Spock, whose mother is human but whose father is Vulcan, has essentially no emotions, giving the story writers wonderful opportunities to pit Spock’s pure reason against Captain Kirk’s human emotions. Similarly, in the later series, Lieutenant Commander Data is pure android, completely artificial, and his lack of emotion provides similar fodder for the writers, although several episodes tinker with the possibility of adding an “emotion chip” into Data, as if emotion were a separate section of the brain that could be added or subtracted at will. But although the series is fiction, the writers did their homework well: their portrayal of the role of emotion in decision making and social interaction is reasonable enough that the psychologists Robert Sekuler and Randolph Blake found them excellent examples of the phenomena, appropriate for teaching introductory psychology. In their book, Star Trek on the Brain, they used numerous examples from the Star Trek series to illustrate the role of emotion in behavior (among other topics).

**Emotional Things**

How will my toaster ever get better, making toast the way I prefer, unless it has some pride? Machines will not be smart and sensible until
they have both intelligence and emotions. Emotions enable us to translate intelligence into action.

Without pride in the quality of our actions, why would we endeavor to do better? The positive emotions are of critical importance to learning, to maintaining our curiosity about the world. Negative emotions may keep us from danger, but it is positive emotions that make living worthwhile, that guide us to the good things in life, that reward our successes, and that make us strive to be better.

Pure reason doesn’t always suffice. What happens when there isn’t enough information? How do we decide which course of action to take when there is risk, so that the possibility of harm has to be balanced against the emotional gain of success? This is where emotions play a critical role and where humans who have had neurological damage to their emotional systems falter. In the movie 2001, the astronaut Dave risks his life in order to recover the dead body of his fellow astronaut. Logically this doesn’t make much sense, but in terms of a long history of human society, it is of great importance. Indeed, this tendency of people to risk many lives in the effort to rescue a few—or even to retrieve the dead—is a constant theme in both our real lives and our fictional ones, in literature, theater, and film.

Robots will need something akin to emotion in order to make these complex decisions. Will that walkway hold the robot’s weight? Is there some danger lurking behind the post? These decisions require going beyond perceptual information to use experience and general knowledge to make inferences about the world and then to use the emotional system to help assess the situation and move toward action. With just pure logic, we could spend all day frozen in place, unable to move as we think through all the possible things that could go wrong—as happens to some emotionally impaired people. To make these decisions we need emotions: robots will, too.

Rich, layered systems of affect akin to that of people are not yet a part of our machines, but some day they will be. Mind you, the affect required is not necessarily a copy of humans’. Instead, what is needed is an affective system tuned to the needs of the system. Robots should be concerned about dangers that might befall them, many of which are common to people and animals, and some of which are unique to robots. They need to avoid falling down stairs or off of other edges, so they should have fear of heights. They should get fatigued, so they do not wear themselves out and become low on energy (hungry?) before recharging their battery. They don’t have to eat or use a toilet, but they do need to be serviced periodically: oil their joints, replace worn parts, and so on. They don’t have to worry about cleanliness and sanitation, but they do need to pay attention to dirt that might get into their moving parts, dust and dirt on their television lenses, and computer viruses that might interfere with their functioning. The affect that robots require will be both similar to and very different from that of people.

Even though machine designers may have never considered that they were building affect or emotion into their machines, they have built in safety and survival systems. Some of these are like the visceral level of people: simple, fast-acting circuits that detect possible danger and react accordingly. In other words, survival has already been a part of most machine design. Many devices have fuses, so that if they suddenly draw more electric current than normal, the fuse or circuit breaker opens the electrical circuit, preventing the machine from damaging itself (and, along the way, preventing it from damaging us or the environment). Similarly, some computers have non-interruptible power supplies, so that if the electric power fails, they quickly and immediately switch to battery power. The battery gives them time to shut down in an orderly and graceful fashion, saving all their data and sending notices to human operators. Some equipment has temperature or water-level sensors. Some detect the presence of people and refuse to operate whenever someone is in a proscribed zone. Existing robots and other mobile systems already have sensors and visual systems that prevent them from hitting people and other objects or falling down stairs. So simple safety and survival is already a part of many designs.

In people and animals, the impact of the visceral system doesn’t cease with an initial response. The visceral level signals higher levels
of processing to try to determine the causes of the problem and to
determine an effective response. Machines should do the same.

Any system that is autonomous—that is intended to exist by itself,
without a caretaker always guiding it—continually has to decide
which of many possible activities to do. In technical terms, it needs a
scheduling system. Even people have difficulty with this task. If we
are working hard to finish an important task, when should we take a
break to eat, sleep, or to do some other activity that is perhaps
required of us but not nearly so urgent? How do we fit the many
activities that have to be done into the limited hours of the day, knowing when to put one aside, when not to? Which is more important:
The critical proposal due tomorrow morning or planning a family
birthday celebration? These are difficult problems that no machine
today can even contemplate but that people face every day. These are
precisely the sorts of decision-making and control problems for which
the emotional system is so helpful.

Many machines are designed to work even though individual com-
ponents may fail. This behavior is critical in safety-related systems,
such as airplanes and nuclear power reactors, and very valuable in
systems that are performing critical operations, such as some computer
systems, hospitals, and anything dealing with the vital infrastruc-
ture of society. But what happens when a component fails and the
automatic backups take over? Here is where the affective system
would be useful.

The component failure should be detected at the visceral level and
used to trigger an alert: in essence, the system would become “anx-
ious.” The result of this increased anxiety should be to cause the
machine to act more conservatively, perhaps slowing down or post-
poning non-critical jobs. In other words, why shouldn’t machines
behave like people who have become anxious? They would be cau-
tious even while attempting to remove the cause of anxiety. With peo-
ple, behavior becomes more focused until the cause and an
appropriate response are determined. Whatever the response for
machine systems, some change in normal behavior is required.

Animals and humans have developed sophisticated mechanisms for
surviving in an unpredictable, dynamic world, coupling the appraisals
and evaluations of affect to methods for modulating the overall sys-
tem. The result is increased robustness and error tolerance. Our arti-
ficial systems would do well to learn from their example.

Emotional Robots

The 1980s was the decade of the PC, the 90s of the Internet, but I believe the
decade just starting will be the decade of the robot.

—Sony Corporation Executive

Suppose we wish to build a robot capable of living in the home, wander-
ing about, fitting comfortably into the family—what would it do? When asked this question, most people first think of handing over
their daily chores. The robot should be a servant, cleaning the house, taking care of the chores. Everyone seems to want a robot that will do
the dishes or the laundry. Actually, today’s dishwashers and clothes
washers and dryers could be considered to be very simple, special-
purpose robots, but what people really have in mind is something that
will go around the house and collect the dirty dishes and clothes, sort
and wash them, and then put them back to their proper places—after,
of course, pressing and folding the clean clothes. All of these tasks are
quite difficult, beyond the capabilities of the first few generations of
robots.

Today, robots are not yet household objects. They show up in science
fairs and factory floors, search-and-rescue missions, and other
specialized events. But this will change. Sony has announced this to be
the decade of the robot, and even if Sony is too optimistic, I do pre-
dict that robots will blossom forth during the first half of the twenty-
first century.
The Young Lady's Illustrated Primer, can take over the entire education of young girls from age four through adulthood. The illustrated primer is still some time in the future, but more limited tutors are already in existence. In addition to education, some robots will do household chores: vacuuming, dusting, and cleaning up. Eventually their range of abilities will expand. Some may end up being built into homes or furniture. Some will be mobile, capable of wandering about on their own.

These developments will require a coevolutionary process of adaptation for both people and devices. This is common with our technologies: we reconfigure the way we live and work to make things possible for our machines to function. The most dramatic coevolution is the automobile system, for which we have altered our homes to include garages and driveways sized and equipped for the automobile, and built a massive worldwide highway system, traffic signaling systems, pedestrian passageways, and huge parking lots. Homes, too, have been transformed to accommodate the multiple wires and pipes of the ever-increasing infrastructure of modern life: hot and cold water, waste return, air vents to the roof, heating and cooling ducts, electricity, telephone, television, internet and home computer and entertainment networks. Doors have to be wide enough for our furniture, and many homes have to accommodate wheelchairs and people using walkers. Just as we have accommodated the home for all these changes, I expect modification to accommodate robots. Slow modification, to be sure, but as robots increase in usefulness, we will ensure their success by minimizing obstacles and, eventually, building charging stations, cleaning and maintenance places, and so on. After all, the vacuum cleaner robot will need a place to empty its dirt, and the garbage robot will need to be able to carry the garbage outside the home. I wouldn't be surprised to see robot quarters in homes, that is, specially built niches where the robots can reside, out of the way, when they are not active. We have closets and pantries for today's appliances, so why not ones especially equipped for robots, with doors that can be controlled by the robot, electrical outlets, interior lights so robots can see to clean

Robots will take many forms. I can imagine a family of robot appliances in the kitchen—refrigerator, pantry, coffeemaker, cooking, and dishwasher robots—all configured to communicate with one another and to transfer food, dishes, and utensils back and forth. The home servant robot wanders about, picking up dirty dishes, delivering them to the dishwasher robot. The dishwasher, in turn, delivers clean dishes and utensils to the robot pantry, which stores them until needed by person or robot. The pantry, refrigerator, and cooking robots work smoothly to prepare the day's menu and, finally, place the completed meal onto dishes provided by the pantry robot.

Some robots will take care of children by playing with them, reading to them, singing songs. Educational toys are already doing this, and the sophisticated robot could act as a powerful tutor, starting with the alphabet, reading, and arithmetic, but soon expanding to almost any topic. Neal Stephenson's science fiction novel, The Diamond Age,
Robots, especially at first, will probably require smooth floors, without obstacles. Door thresholds might have to be eliminated or minimized. Some locations—especially stairways—might have to be especially marked, perhaps with lights, infrared transmitters, or simply special reflective tape. Barcodes or distinctive markers posted here and there in the home would enormously simplify the robot’s ability to recognize its location.

Consider how a servant robot might bring a drink to its owner. Ask for a can of soda, and off goes the robot, obediently making its way to the kitchen and the refrigerator, which is where the soda is kept. Understanding the command and navigating to the refrigerator are relatively simple. Figuring out how to open the door, find the can, and extract it is not so simple. Giving the servant robot the dexterity, the strength, and the non-slip wheels that would allow it to pull open the refrigerator door is quite a feat. Providing the vision system that can find the soda, especially if it is completely hidden behind other food items, is difficult, and then figuring out how to extract the can without destroying objects in the way is beyond today’s capabilities in robot arms.

How much simpler it would be if there were a drink dispenser robot tailored to the needs of the servant robot. Imagine a drink-dispensing robot appliance capable of holding six or twelve cans, refrigerated, with an automatic door and a push-arm. The servant robot could go to the drink robot, announce its presence and its request (probably by an infrared or radio signal), and place its tray in front of the dispenser. The drink robot would slide open its door, push out a can, and close the door again: no complex vision, no dexterous arm, no forceful opening of the door. The servant robot would receive the can on its tray, and then go back to its owner.

In a similar way, we might modify the dishwasher to make it easier for a home robot to load it with dirty dishes, perhaps give it special trays with designated slots for different dishes. But as long as we are doing that, why not make the pantry a specialized robot, one capable of removing the clean dishes from the dishwasher and storing them for later use? The special trays would help the pantry as well. Perhaps the pantry could automatically deliver cups to the coffeemaker and plates to the home cooking robot, which is, of course, connected to refrigerator, sink, and trash. Does this sound far-fetched? Perhaps, but, in fact, our household appliances are already complex, many of them with multiple connections to services. The refrigerator has connections to electric power and water. Some are already connected to the internet. The dishwasher and clothes washer have electricity, water and sewer connections. Integrating these units so that they can work smoothly with one another does not seem all that difficult.

I imagine that the home will contain a number of specialized robots: the servant is perhaps the most general purpose, but it would work together with a cleaning robot, the drink dispensing robot, perhaps some outside gardening robots, and a family of kitchen robots, such as dishwasher, coffee-making, and pantry robots. As these robots are developed, we will probably also design specialized objects in the home that simplify the tasks for the robots, coevolving robot and home to work smoothly together. Note that the end result will be better for people as well. Thus, the drink dispenser robot would allow anyone to walk up to it and ask for a can, except that you wouldn’t use infrared or radio, you might push a button or perhaps just ask.

I am not alone in imagining this coevolution of robots and homes. Rodney Brooks, one of the world’s leading roboticists, head of the MIT Artificial Intelligence Laboratory and founder of a company that builds home and commercial robots, imagines a rich ecology of environments and robots, with specialized ones living on devices, each responsible to keep its domain clean: one does the bathtub, another the toilet; one does windows, another manipulates mirrors. Brooks even contemplates a robot dining room table, with storage area and dishwasher built into its base so that “when we want to set the table, small robotic arms, not unlike the ones in a jukebox, will bring the required dishes and cutlery out onto the place settings. As each course is fin-
ished, the table and its little robot arms would grab the plates and devour them into the large internal volume underneath.”

What should a robot look like? Robots in the movies often look like people, with two legs, two arms, and a head. But why? Form should follow function. The fact that we have legs allows us to navigate irregular terrain, something an animal on wheels could not do. The fact that we have two hands allows us to lift and manipulate, with one hand helping the other. The humanoid shape has evolved over eons of interaction with the world to cope efficiently and effectively with it. So, where the demands upon a robot are similar to those upon people, having a similar shape might be sensible.

If robots don’t have to move—such as drink, dishwasher, or pantry robots—they need not have any means of locomotion, neither legs nor wheels. If the robot is a coffeemaker, it should look like a coffeemaker, modified to allow it to connect to the dishwasher and pantry. If the robot is a lawn mower, it should look like a lawn mower. If the robot is a vacuum cleaner, it should look like a vacuum cleaner. If the robot is a lawn mower that connects to the dishwasher and pantry, it should look like a lawn mower that connects to the dishwasher and pantry.

Figure 6.3
What should a robot look like?
The Roomba is a vacuum cleaner, its shape appropriate for running around the floor and maneuvering itself under the furniture. This robot doesn’t look like either a person or an animal, nor should it: its shape fits the task.
(Courtesy of iRobot Inc.)

body to hold the batteries and to support the legs, wheels, or tracks for locomotion; hands to pick up objects; and cameras (eyes) on top where they can better survey the environment. In other words, some robots will look like an animal or human, not because this is cute, but because it is the most effective configuration for the task. These robots will probably look something like R2D2 (figure 6.1): a cylindrical or rectangular body on top of some wheels, tracks, or legs; some form of manipulable arm or tray; and sensors all around to detect obstacles, stairs, people, pets, other robots, and, of course the objects they are supposed to interact with. Except for pure entertainment value, it is difficult to understand why we would ever want a robot that looked like C3PO.

In fact, making a robot humanlike might backfire, making it less acceptable. Masahiro Mori, a Japanese roboticist, has argued that we are least accepting of creatures that look very human, but that per-
Affect and Emotion in Robots

What emotions will a robot need to have? The answer depends upon the sort of robot we are thinking about, the tasks it is to perform, the nature of the environment, and what its social life is like. Does it interact with other robots, animals, machines, or people? If so, it will need to express its own emotional state as well as to assess the emotions of the people and animals it interacts with.

Think about the average, everyday home robot. These don’t yet exist, but some day the house will become populated with robots. Some home robots will be fixed in place, specialized, such as kitchen robots: for example, the pantry, dishwasher, drink dispenser, food dispenser, coffeemaker, or cooking unit robots. And, of course, clothes washer, drier, iron, and clothes-folding robots, perhaps coupled to wardrobe robots. Some will be mobile, but also specialized, such as the robots that vacuum the floors and mow the lawn. But probably we will also have at least one general-purpose robot: the home servant robot, that brings us coffee, cleans up, does simple errands, and looks after and supervises the other robots. It is the home robot that is of most interest, because it will have to be the most flexible and advanced.

Servant robots will need to interact with us and with the other robots of the house. For the other robots, they could use wireless communication. They could discuss the jobs they were doing, whether or not they were overloaded or idle. They could also state when they were running low on supplies and when they sensed difficulties, problems, or errors and call upon one another for help. But what about when robots interact with people? How will this happen?

Servant robots need to be able to communicate with their owners. Some way of issuing commands is needed, some way of clarifying the ambiguities, changing a command in midstream (“Forget the coffee, bring me a glass of water instead”), and dealing with all of the complexities of human language. Today, we can’t do that, so robots that are built now will have to rely upon very simple commands or even some sort of remote controller, where a person pushes the appropriate buttons, generates a well-structured command, or selects actions from a menu. But the time will come when we can interact in speech, with the robots understanding not just the words but the meanings behind them.

When should a robot volunteer to help its owners? Here, robots will need to be able to assess the emotional state of people. Is someone struggling to do a task? The robot might want to volunteer to help. Are the people in the house arguing? The robot might wish to go to some other room, out of the way. Did something bring pleasure? The robot might wish to remember that, so it could do it again when appropriate. Was an action poorly done, so the person showed disappointment? Perhaps the action could be improved, so that next time the robot
would produce better results. For all these reasons, and more, the robot will need to be designed with the ability to read the emotional state of its owners.

A robot will need to have eyes and ears (cameras and microphones) to read facial expressions, body language, and the emotional components of speech. It will have to be sensitive to tones of voice, the tempo of speech, and its amplitude, so that it can recognize anger, delight, frustration, or joy. It needs to be able to recognize scolding voices from praising ones. Note that all of these states can be recognized just by their sound quality without the need to recognize the words or language. Notice that you can determine other people’s emotional states just by the tone of voice alone. Try it: Make believe you are in any one of those states—angry, happy, scolding, or praising—and express yourself while keeping your lips firmly sealed. You can do it entirely with the sounds, without speaking a word. These are universal sound patterns.

Similarly, the robot should display its emotional state, much as a person does (or, perhaps more appropriately, as a pet dog or child does), so that the people with whom it is interacting can tell when a request is understood, when it is something easy to do, difficult to do, or perhaps even when the robot judges it to be inappropriate. Similarly, the robot should show pleasure and displeasure, an energetic appearance or exhaustion, confidence or anxiety when appropriate. If it is stuck, unable to complete a task, it should show its frustration. It will be as valuable for the robot to display its emotional state as it is for people to do so. The expressions of the robot will allow us humans to understand the state of the robot, thereby learning which tasks are appropriate for it, which are not. As a result, we can clarify instructions or even offer help, eventually learning to take better advantage of the robot’s capabilities.

Many people in the robotics and computer research community believe that the way to display emotions is to have a robot decide whether it is happy or sad, angry or upset, and then display the appropriate face, usually an exaggerated parody of a person in those states. I argue strongly against this approach. It is fake, and, moreover, it looks fake. This is not how people operate. We don’t decide that we are happy, and then put on a happy face, at least not normally. This is what we do when we are trying to fool someone. But think about all those professionals who are forced to smile no matter what the circumstance: they fool no one—they look just like they are forcing a smile, as indeed they are.

The way humans show facial expression is by automatic innervation of the large number of muscles involved in controlling the face and body. Positive affect leads to relaxation of some muscle groups, automatic pulling up of many facial muscles (hence the smile, raised eyebrows and cheeks, etc.), and a tendency to open up and draw closer to the positive event or thing. Negative affect has the opposite impact, causing withdrawal, to push away. Some muscles are tensed, and some of the facial muscles pull downward (hence the frown). Most affective states are complex mixtures of positive and negative valence, at differing levels of arousal, with some residue of the immediately previous states. The resulting expressions are rich and informative. And real.

Fake emotions look fake: we are very good at detecting false attempts to manipulate us. Thus, many of the computer systems we interact with—the ones with cute, smiling helpers and artificially sweet voices and expressions—tend to be more irritating than useful. “How do I turn this off?” is a question often asked of me, and I have become adept at disabling them, both in my own computers or those of others who seek to be released from the irritation.

I have argued that machines should indeed both have and display emotions, the better for us to interact with them. This is precisely why the emotions need to appear as natural and ordinary as human emotions. They must be real, a direct reflection of the internal states and processing of a robot. We need to know when a robot is confident or confused, secure or worried, understanding our queries or not, working on our request or ignoring us. If the facial and body expressions reflect the underlying processing, then the emotional displays will
seem genuine precisely because they are real. Then we can interpret their state, they can interpret ours, and the communication and interaction will flow ever more smoothly.

I am not the only person to have reached this conclusion. MIT Professor Rosalind Picard once said, talking about whether robots should have emotions, “I wasn’t sure they had to have emotions until I was writing up a paper on how they would respond intelligently to our emotions without having their own. In the course of writing that paper, I realized it would be a heck of a lot easier if we just gave them emotions.”

Once robots have emotions, then they need to be able to display them in a way that people can interpret—that is, as body language and facial expressions similar to human ones. Thus, the robot’s face and body should have internal actuators that act and react like human muscles according to the internal states of the robot. People’s faces are richly endowed with muscle groups in chin, lips, nostrils, eyebrows, forehead, cheeks, and so on. This complex of muscles makes for a sophisticated signaling system, and if robots were created in a similar way, the features of the face will naturally smile when things are going well and frown when difficulties arise. For this purpose, robot designers need to study and understand the complex workings of human expressions, with its very rich set of muscles and ligaments tightly intertwined with the affective system.

Displaying full facial emotions is actually very difficult. Figure 6.4 shows Leonardo, Professor Cynthia Breazeal’s robot at the MIT Media Laboratory, designed to control a vast array of facial features, neck, body, and arm movements, all the better to interact socially and emotionally with us. There is a lot going on inside our bodies, and much the same complexity is required within the faces of robots.

But what of the underlying emotional states? What should these be? As I’ve discussed, at the least, the robot should be cautious of heights, wary of hot objects, and sensitive to situations that might lead to hurt or injury. Fear, anxiety, pain, and unhappiness might all be appropriate states for a robot. Similarly, it should have positive states, including pleasure, satisfaction, gratitude, happiness and pride, which would enable it to learn from its actions, to repeat the positive ones and improve, where possible.

Surprise is probably essential. When what happens is not what is expected, the surprised robot should interpret this as a warning. If a room unexpectedly gets dark, or maybe the robot bumps into something it didn’t expect, a prudent response is to stop all movement and figure out why. Surprise means that a situation is not as anticipated, and that planned or current behavior is probably no longer appropriate—hence, the need to stop and reassess.

Some states, such as fatigue, pain, or hunger, are simpler, for they do not require expectations or predictions, but rather simple monitoring of internal sensors. (Fatigue and hunger are technically not affective states, but they can be treated as if they were.) In the human,
The servant robot should get frustrated waiting for the coffee, so it should temporarily give up. As soon as the servant robot gives up the quest for coffee, it is free to attend to the dishwasher’s request, go off and find the dirty coffee cups. This would automatically solve the deadlock: the servant robot would find some dirty cups, deliver them to the dishwasher, which would eventually let the coffeemaker make the coffee and let me get my coffee, although with some delay.

Could the servant robot learn from this experience? It should add to its list of activities the periodic collection of dirty dishes, so that the dishwasher/pantry would never run out again. This is where some pride would come in handy. Without pride, the robot doesn’t care: it has no incentive to learn to do things better. Ideally, the robot would take pride in avoiding difficulties, in never getting stuck at the same problem more than once. This attitude requires that robots have positive emotions, emotions that make them feel good about themselves, that cause them to get better and better at their jobs, to improve, perhaps even to volunteer to do new tasks, to learn new ways of doing things. Pride in doing a good job, in pleasing their owners.

Machines That Sense Emotion

The extent to which emotional upsets can interfere with mental life is no news to teachers. Students who are anxious, angry, or depressed don’t learn; people who are caught in these states do not take in information efficiently or deal with it well.

—Daniel Goleman, Emotional Intelligence

Suppose machines could sense the emotions of people. What if they were as sensitive to the moods of their users as a good therapist might be? What if an electronic, computer-controlled educational system...
Figure 6.5
MIT's Affective Computing program.
The diagram indicates the complexity of the human affective system and the challenges required to monitor affect properly. From the work of Prof. Rosalind Picard of MIT.
(Drawing courtesy of Roz Picard and Jonathan Klein.)

could sense when the learner was doing well, was frustrated, or was proceeding appropriately? Or what if the home appliances and robots of the future could change their operations according to the moods of their owners? What then?

Professor Rosalind Picard at the MIT Media Laboratory leads a research effort entitled “Affective Computing,” an attempt to develop machines that can sense the emotions of the people with whom they are interacting, and then respond accordingly. Her research group has made considerable progress in developing measuring devices to sense fear and anxiety, unhappiness and distress. And, of course, satisfaction

and happiness. Figure 6.5 is taken from their web site and demonstrates the variety of issues that must be addressed.

How are someone’s emotions sensed? The body displays its emotional state in a variety of ways. There are, of course, facial expressions and body language. Can people control their expressions? Well, yes, but the visceral layer works automatically, and although the behavioral and reflective levels can try to inhibit visceral reaction, complete suppression does not appear to be possible. Even the most controlled person, the so-called poker-face who keeps a neutral display of emotional responses no matter what the situation, still has micro-expressions—short, fleeting expressions that can be detected by trained observers.

In addition to the responses of one’s musculature, there are many physiological responses. For example, although the size of the eye’s pupil is affected by light intensity, it is also an indicator of emotional arousal. Become interested or emotionally aroused, and the pupil widens. Work hard on a problem, and it widens. These responses are involuntary, so it is difficult—probably impossible—for a person to control them. One reason professional gamblers sometimes wear tinted eyeglasses even in dark rooms is to prevent their opponents from detecting changes in the size of their pupils.

Heart rate, blood pressure, breathing rate, and sweating are common measures that are used to derive affective state. Even amounts of sweating so small that the person can be unaware of it can trigger a change in the skin’s electrical conductivity. All of these measures can readily be detected by the appropriate electronics.

The problem is that these simple physiological measures are indirect measures of affect. Each is affected by numerous things, not just by affect or emotion. As a result, although these measures are used in many clinical and applied settings, they must be interpreted with care. Thus, consider the workings of the so-called lie detector. A lie detector is, if anything, an emotion detector. The method is technically called “polygraph testing” because it works by simultaneously record-
ing and graphing multiple physiological measures such as heart rate, breathing rate, and skin conductance. A lie detector does not detect falsehoods; it detects a person’s affective response to a series of questions being asked by the examiner, where some of the answers are assumed to be truthful (and thus show low affective responses) and some deceitful (and thus show high affective arousal). It is easy to see why lie detectors are so controversial. Innocent people might have large emotional responses to critical questions while guilty people might show no response to the same questions.

Skilled operators of lie detectors try to compensate for these difficulties by the use of control questions to calibrate a person’s responses. For example, by asking a question to which they expect a lie in response, but that is not relevant to the issue at hand, they can see what a lie response looks like in the person being tested. This is done by interviewing the suspect and then developing a series of questions designed to ferret out normal deviant behavior, behavior in which the examiner has no interest, but where the suspect is likely to lie. One question commonly used in the United States is “Did you ever steal something when you were a teenager?”

Because lie detectors record underlying physiological states associated with emotions rather than with lies, they are not very reliable, yielding both misses (when a lie is not detected because it produces no emotional response) and false alarms (when the nervous suspect produces emotional responses even though he or she is not guilty). Skilled operators of these machines are aware of the pitfalls, and some use the lie detector test as a means of eliciting a confession: people who truly believe the lie detector can “read minds” might confess just because of their fear of the test. I have spoken to skilled operators who readily agree to the critique I just provided, but are proud of their record of eliciting voluntary confessions. But even innocent people have sometimes confessed to crimes they did not commit, strange as this might seem. The record of accuracy is flawed enough that the National Research Council of the United States National Academies performed a lengthy, thorough study and concluded that polygraph testing is too flawed for security screening and legal use.

Suppose we could detect a person’s emotional state, then what? How should we respond? This is a major, unsolved problem. Consider the classroom situation. If a student is frustrated, should we try to remove the frustration, or is the frustration a necessary part of learning? If an automobile driver is tense and stressed, what is the appropriate response?

The proper response to an emotion clearly depends upon the situation. If a student is frustrated because the information provided is not clear or intelligible, then knowing about the frustration is important to the instructor, who presumably can correct the problem through further explanation. (In my experience, however, this often fails, because an instructor who causes such frustration in the first place is usually poorly equipped to understand how to remedy the problem.)

If the frustration is due to the complexity of the problem, then the proper response of a teacher might be to do nothing. It is normal and proper for students to become frustrated when attempting to solve problems slightly beyond their ability, or to do something that has never been done before. In fact, if students aren’t occasionally frustrated, it probably is a bad thing—it means they aren’t taking enough risks, they aren’t pushing themselves sufficiently.

Still, it probably is good to reassure frustrated students, to explain that some amount of frustration is appropriate and even necessary. This is a good kind of frustration that leads to improvement and learning. If it goes on too long, however, the frustration can lead students to give up, to decide that the problem is above their ability. Here is where it is necessary to offer advice, tutorial explanations, or other guidance.

What of frustrations shown by students that have nothing to do with the class, that might be the result of some personal experience, outside the classroom? Here it isn’t clear what to do. The instructor,
whether person or machine, is not apt to be a good therapist. Expressing sympathy might or might not be the best or most appropriate response.

Machines that can sense emotions are an emerging new frontier of research, one that raises as many questions as it addresses, both in how machines might detect emotions and in how to determine the most appropriate way of responding. Note that while we struggle to determine how to make machines respond appropriately to signs of emotions, people aren’t particularly good at it either. Many people have great difficulty responding appropriately to others who are experiencing emotional distress: sometimes their attempts to be helpful make the problem worse. And many are surprisingly insensitive to the emotional states of others, even people whom they know well. It is natural for people under emotional strain to try to hide the fact, and most people are not experts in detecting emotional signs.

Still, this is an important research area. Even if we are never able to develop machines that can respond completely appropriately, the research should inform us both about human emotion and also about human-machine interaction.

Machines That Induce Emotion in People

It is surprisingly easy to get people to have an intense emotional experience with even the simplest of computer systems. Perhaps the earliest such experience was with Eliza, a computer program developed by the MIT computer scientist Joseph Weizenbaum. Eliza was a simple program that worked by following a small number of conversational scripts that had been prepared in advance by the programmer (originally, this was Weizenbaum). By following these scripts, Eliza could interact with a person on whatever subject the script had prepared it for. Here is an example. When you started the program, it would greet you by saying: “Hello. I am ELIZA. How can I help you?” If you responded by typing: “I am concerned about the increasing level of violence in the world,” Eliza would respond: “How long have you been concerned about the increasing level of violence in the world?” That’s a relevant question, so a natural reply would be something like, “Just the last few months,” to which Eliza would respond, “Please go on.”

You can see how you might get captured by the conversation: your concerns received sympathetic responses. But Eliza has no understanding of language. It simply finds patterns and responds appropriately (saying “Please go on” when it doesn’t recognize the pattern). Thus, it is easy to fool Eliza by typing: “I am concerned about abc, def, and for that matter, ghi,” to which Eliza would dutifully reply: “How long have you been concerned about abc, def, and for that matter, ghi?”

Eliza simply recognizes the phrase “I am concerned about X” and replies, “How long have you been concerned about X?” with absolutely no understanding of the words.

Because most people took Eliza seriously, they did not try to trick it. Instead, they took each reply seriously, pondering their hidden meanings. As a result, they would sometimes discover themselves in deep, philosophical debate or, in the most popular version of Eliza that was scripted to act like a psychotherapist, they would find themselves discussing their most intimate secrets.

In the mid-1960s, one of my long-term friends and research collaborators, Daniel Bobrow, was a research scientist at BBN, a Cambridge, Massachusetts company doing research in artificial intelligence, computer networks, and acoustics. Bobrow told me that a programmer, having worked hard to get Eliza working on a BBN computer, finally finished about 5:00 A.M. Saturday morning and then went home, leaving the program running on the computer. (In those days—this was 1966—personal computers were unheard of, and only advanced companies, such as BBN, even had computers for its employees to use, which is one reason programmers had to work at night to get computer time.)

The vice president of Bobrow’s division came in later that Saturday morning to demonstrate the computer to some customers. Seeing that
As Bobrow told me when I discussed this interaction with him, "You can see he cared a lot about the answers to his questions, and what he thought were my remarks had an emotional effect on him." We are extremely trusting, which makes us very easy to fool, and very angry when we think we aren't being taken seriously.

The reason Eliza had such a powerful impact is related to the discussions in chapter 5 on the human tendency to believe that any intelligent-seeming interaction must be due to a human or, at least, an intelligent presence: anthropomorphism. Moreover, because we are trusting, we tend to take these interactions seriously. Eliza was written a long time ago, but its creator, Joseph Weizenbaum, was horrified by the seriousness with which his simple system was taken by so many people who interacted with it. His concerns led him to write Computer Power and Human Reason, in which he argued most cogently that these shallow interactions were detrimental to human society.

We have come a long way since Eliza was written. Computers of today are thousands of times more powerful than they were in the 1960s and, more importantly, our knowledge of human behavior and psychology has improved dramatically. As a result, today we can write programs and build machines that, unlike Eliza, have some true understanding and can exhibit true emotions. However, this doesn’t mean that we have escaped from Weizenbaum’s concerns. Consider Kismet.

Kismet, whose photograph is shown in figure 6.6, was developed by a team of researchers at the MIT Artificial Intelligence Laboratory and reported upon in detail in Cynthia Breazeal’s Designing Sociable Robots.

Recall that the underlying emotions of speech can be detected without any language understanding. Angry, scolding, pleading, consoling, grateful, and praising voices all have distinctive pitch and loudness contours. We can tell which of these states someone is in even if they are speaking in a foreign language. Our pets can often detect our moods through both our body language and the emotional patterns within our voices.

Kismet uses these cues to detect the emotional state of the person...
with whom it is interacting. Kismet has video cameras for eyes and a microphone with which to listen. Kismet has a sophisticated structure for interpreting, evaluating, and responding to the world—shown in figure 6.7—that combines perception, emotion, and attention to control behavior. Walk up to Kismet, and it turns to face you, looking you straight in the eyes. But if you just stand there and do nothing else, Kismet gets bored and looks around. If you do speak, it is sensitive to the emotional tone of the voice, reacting with interest and pleasure to encouraging, rewarding praise and with shame and sorrow to scolding. Kismet’s emotional space is quite rich, and it can move its head, neck, eyes, ears, and mouth to express emotions. Make it sad, and its ears droop. Make it excited and it perks up. When unhappy, the head droops, ears sag, mouth turns down.

Interacting with Kismet is a rich, engaging experience. It is difficult to believe that Kismet is all emotion, with no understanding. But walk up to it, speak excitedly, show it your brand-new watch, and Kismet responds appropriately: it looks at your face, then at the watch, then back at your face again, all the time showing interest by raising its eyelids and ears, and exhibiting perky, lively behavior. Just the interested responses you want from your conversational partner, even though Kismet has absolutely no understanding of language or, for that matter, your watch. How does it know to look at the watch? It doesn’t, but it responds to movement so it looks at your rising hand. When the motion stops, it gets bored, and returns to look at your eyes. It shows excitement because it detected the tone of your voice.

Note that Kismet shares some characteristics of Eliza. Thus, although this is a complex system, with a body (well, a head and neck), multiple motors that serve as muscles, and a complex underlying model of attention and emotion, it still lacks any true understand-
ing. Therefore, the interest and boredom that it shows toward people are simply programmed responses to changes—or the lack thereof—in the environment and responses to movement and physical aspects of speech. Although Kismet can sometimes keep people entranced for long periods, the enhancement is somewhat akin to that of Eliza: most of the sophistication is in the observer’s interpretations.

Aibo, the Sony robot dog, has a far less sophisticated emotional repertoire and intelligence than Kismet. Nonetheless, Aibo has also proven to be incredibly engaging to its owners. Many owners of the robot dog band together to form clubs: some own several robots. They trade stories about how they have trained Aibo to do various tricks. They share ideas and techniques. Some firmly believe that their personal Aibo recognizes them and obeys commands even though it is not capable of these deeds.

When machines display emotions, they provide a rich and satisfying interaction with people, even though most of the richness and satisfaction, most of the interpretation and understanding, comes from within the head of the person, not from the artificial system. Sherry Turkle, both an MIT professor and a psychoanalyst, has summarized these interactions by pointing out, “It tells you more about us as human beings than it does the robots.” Anthropomorphism again: we read emotions and intentions into all sorts of things. “These things push on our buttons whether or not they have consciousness or intelligence,” Turkle said. “They push on our buttons to recognize them as though they do. We are programmed to respond in a caring way to these new kinds of creatures. The key is these objects want you to nurture them and they thrive when you pay attention.”