

Affective Issues in the Search for Artificial Intelligence

By Jason Aughenbaugh

Jma@princeton.edu

Psychology 322: Human Machine Interactions
Final Project

Advisor: Professor J. J. Gelfand

January 9, 1999

Recent studies in robotics, artificial intelligence, and the emerging field of affective computing have roots in the most basic human qualities. The human ability to grip and manipulate tools has played a crucial role in humankind's ascent to a dominant position on Earth. The introductions of fire, stone, metal, industry, electricity, and computers into human lives have each had profound effects on human lifestyles. In the absence of the most basic tools, such as spears and rocks, humans might not have survived the natural selection process. Modern researchers seek to further enlarge the human toolset. While it took thousands of years to progress from a spear to a rifle, automobiles, airplanes, and space vehicles all developed in the last one hundred years. Information systems are an even more recent development. The Internet and personal computer only became household words in the last 15 years. Without elaborate communication (languages) and social interaction, there would be no pyramids, aqueducts, cars or computers. Many would argue that there would be little to human life without the arts and humanities, which have strong roots in human emotions and feelings. Emotions have played as large role in our evolution as intellect and brawn, both physically and culturally. However, computers, information, and intelligence heretofore have been separated from emotions. Now they are being combined in ways that will shape the twenty-first century.

Henry Ford is generally credited with the first successful use of an assembly line. This was in many ways an early step towards automation. Distinct units (humans) had specific jobs to do. Once the specific job was learned, it could be done well—efficiently and consistently. However, human workers sometimes grew tired and bored of their monotonous and sometimes dangerous jobs. Machines were developed to replace human workers at these tasks, and factories started to become automated. Much as a wrench strengthened the human hand by replacing the fingers, machines began to replace the arm. Recently, machines have assumed functions previously reserved to the human brain. These specialized machines are merely turned on and then proceed to cut, solder, bolt, and rivet without human intervention. Automation has become so advanced and widespread today that some factories require only a few humans for operation; the machines are capable of running themselves.

The computer is a specialized machine. Currently, its specialization is computations, calculations, and other number crunching activities. Underlying word processors, browsers, and computer games are numbers. Efforts have been made to create intelligent computers, but such endeavors into artificial intelligence, or AI, have until recently only yielded computers with specialized capabilities. These expert systems are preprogrammed to do certain tasks under certain conditions. Much as a machine on a car assembly line would not be of much use if it were given a table to work on, the majority of AI products are not adaptable. IBM's Deep Blue is an excellent chess player, but it likely would not fair well at checkers, despite the logical nature of both games. In order to learn to play checkers, Deep Blue would have to be completely reprogrammed by humans, if not rebuilt. It probably would not take Garry Kasparov as long to learn checkers.

Recent efforts in AI have begun to examine more generalized applications. Two important ones are machines that can learn from their environment and machines that interact well with people. The new field of affective computing has developed as a result of these endeavors, and it has been fueled by recent findings that emotions play an essential role in rational thought in humans (Damasio, 1994). Affective computing refers to “computing that relates to, arises from, or deliberately influences emotions” (Picard, 1997, p. 3). This involves implementing emotions in machines and creating machines that can recognize and respond to human emotions. One final goal of affective computing is to create a machine that has emotions that it can regulate, reason about, and use intelligently to act rationally in direct social interaction with humans. The general goal of researchers is to achieve an affective performance equal to humans. Accordingly, Picard suggests a revised Turing Test (Turing 1950) to determine the emotional and social intelligence of these systems (Picard, 1997, pp. 12-13).

It may be shocking to speak about emotions in the same breath as computers. The traditional view of computers as computational machines prevails not only in general society, but it is embraced by much of the artificial intelligence community. Because affective computing is the newest branch of AI, detailed discussion of affective systems must begin with the foundations of AI.

Artificial intelligence has long been a dream of many in the computer community. It certainly has consumed the minds of many science fiction authors. An early push for an artificially intelligent system came in 1956 by the CIA. It was interested in building a system that could automatically translate documents from foreign languages into English (Trull, 1998). The original belief was that this could be achieved by using some specifiable, deterministic grammar and syntax. Therefore, by following these inflexible rules, a computer could infallibly translate. Similarly, most early AI endeavors focused on using systems to manipulate symbols and numbers according to rigid rules. Strong AI preserves this interpretation today. Strong AI ensures that all intelligence results from application of various algorithms. In this view, the human brain is a very sophisticated algorithm processor.

Opponents to Strong AI point to several different arguments. The work of David Hilbert, Kurt Godel, and Alan Turing has culminated in the conclusion that algorithms cannot encompass all of mathematics (Trull, 1998). Therefore, it seems unlikely that all of intelligence could rest on algorithms alone. A second argument is best embodied by John Searle's Chinese room argument (Searle, 1990; Searle, 1984). Searle argues that syntax alone is not sufficient for intelligence; semantics are required. A machine (such as a computer) has no way to get from the symbols to the semantics. Picard (1997, p. 12) notes that Johnson-Laird and Shafir have written about the failure of formal logic to determine which of an infinite number of conclusions are sensible to draw from a given set of premises (Johnson-Laird and Shafir, 1993). Another argument stems from Damasio's research into the role of emotions in human intelligence. By studying humans with prefrontal cortex damage, he concludes that emotions are a requisite part of rational thought (Damasio, 1994). Since computers presently have no implementation of emotions, they cannot be truly intelligent; they cannot demonstrate human intelligence.

The battle between emotions and reason has been a recurring theme, not only in the history of AI, but in the history of human civilization (Newell). The long, distinct separation of the two is etched in our use of words “heart” and “head” to refer to emotions and reason, respectively. This deep-rooted distinction causes discomfort with the recent evidence of the role of emotions in reason. Rigid stereotypes also affect the perception of emotions, at least in American culture. The word “emotional” is linked with weakness and irrationality. Certainly, too much emotion, especially unchecked displays of it, can be damaging. This leads to a concept of emotional intelligence that refers to interpersonal and intrapersonal skills (Picard, 1997, p. 13). Peter Salovey and John Mayer describe emotional intelligence as “the ability to monitor one's own and others' feelings and emotions, to discriminate among them, and to use this information to guide one's thinking and actions” (referenced in Picard, 1997, p. 13). It is within this framework that emotions become a part of artificial intelligence. Woody Bledsoe described his AI dream in his 1985 American Association of Artificial Intelligence Presidential Address. He spoke of the “ ‘excitement of seeing a machine act like a human being, at least in many ways’ of building a machine that could ‘understand, act autonomously, think, learn, enjoy, hate’ and which ‘liked to walk and play Ping-Pong, especially with me’ ” (Bates, 1994; Mateas). To meet these criteria, a machine needs a “head” *and* a “heart.”

Another characteristic of earlier AI projects was a passive interaction with their environments. For example, modeling and operations research applications are generally preprogrammed with all parameters and rules. While they might be able to actively ask for more inputs and return error messages and results, these systems do not actively seek information, and they certainly do not get any directly from their environments. Originally, this was primarily a consequence of limited processing power and other technology. New technology and psychological evidence have combined to produce a new interest in creating systems that actively interact with their environment. These devices examine reality—that is the actual situation they are in—and then make decisions and act based on their observations. Even if they are designed to perform one task well, they still have generalized abilities that allow them to adapt and react to the unexpected. Computers that can interact socially or physically with their environments have already been developed. One early example of social interaction was Eliza, a program written by J. Weizenbaum that “conversed” with the user while imitating a Rogerian psychotherapist. Uses for active interaction in a social context will be discussed later. However, social interaction has only recently begun to receive significant attention.

Active interaction in a physical context receives much more attention in society, and in AI. Machines that successfully interact with their physical environment frequently grace news reports. For example, Sojourner, the recent NASA Mars probe, and Dante, a NASA project to explore active volcano projects, are both rovers that must navigate unmapped terrain and adjust to physical changes in their environments (*Robots Rising*, 1996). They must be able to react without constant human contact, and even a slight failure could be fatal. These types of interactions are at the heart of current robotics research, as well as much science fiction.

Robots can be divided into three categories based on control (*Robots Rising*, 1996). Remote control is the lowest level. Remotely controlled robots are not very different from toy remote control cars. A human user directly provides all actions of remotely controlled robots. The next level is supervisory control. At this level, there is no constant guidance from an outside source. However, there is some guidance, especially when unpredicted events occur. The final and most sought after level is autonomy—robots that need no external control.

Another classification of robots is made by Rodney Brooks. He refers to the qualities “embodied” and “situated” (Brooks, 1991). He characterizes situatedness robots that “are situated in the world—they do not deal with abstract descriptions, but with the 'here' and 'now' of the environment that directly influences the behavior of the system.” Embodiment is characterized by robots that “have bodies and experience the world directly--their actions are part of a dynamic with the world, and the actions have immediate feedback on the robot's own sensations.” He continues:

An airline reservation system is situated but it is not embodied--it deals with thousands of requests per second, and its responses vary as its database changes, but it interacts with the world only through sending and receiving messages. A current generation industrial spray-painting robot is embodied but it is not situated--it has physical extent and its servo routines must correct for its interactions with gravity and the noise present in the system, but it does not perceive any aspects of the shape of an object presented to it for painted and simply goes through a pre-programmed series of actions.

The challenge is to create systems that are situated, embodied, and autonomous. These efforts are generally undertaken in the field of robotics.

Robots, especially humanoids (robots with human-like features and capable of human interactions), are often at the heart of artificial intelligence dreams and ideas. Many scientists, children, and movie fans would love a C3-PO (a reference to a popular humanoid in George Lucas's *StarWars* films) of their own. Woody Bledsoe was obviously describing a humanoid in his AI dream. This popularity of humanoids, as well as their status as a holy grail for AI, drives many researchers to work towards creating one. Consequently, they try to model intelligence from the top down, as in top of the food chain; they attempt to model adult human intelligence. This may seem logical since, to the best of our knowledge, humans are the most sophisticated and intelligent creatures. Certainly, they are the most intelligent that we can study. Nevertheless, this may not be the best approach. By trying to begin with the most complex creatures, researchers are forced to create simplified worlds in which they can interact. The robots cannot interact successfully with the real world. D. Marr called in 1976 for examining the simpler components of intelligence, insisting that the efforts of his time were wrong (Marr, 1976).

Rodney Brooks agrees on the importance of the simpler components of intelligence. He has explored two different approaches at MIT. In the mid-80s, Brooks boldly asked the world to take a hint from evolution (Brooks, 1986). Paraphrasing, he notes that single cell life began around 3.5 billion years ago. Mammals developed only

250 million years ago. Immediate predecessors to the great apes developed just 18 million years ago, and humans a mere 2.5 million years ago. Writing was developed less than 5000 years ago, and “expert” knowledge only over the last several hundred years. Then he daringly concludes, “This suggests that problem solving behavior, language, expert knowledge and application, reason, etc., are all pretty simple once the essence of being and reacting are available.”

Brooks proposes to approach artificial intelligence from the bottom up, starting with insects.

Insects are not usually thought of as intelligent. However, they are very robust devices. They operate in a dynamic world, carry out a number of complex tasks including hunting, eating, mating, nest building, and rearing of young. There may be rain, strong winds, predators, and variable food supplies all of which impair the insects' abilities to achieve its goals. Statistically, however, insects succeed. No human-built systems are remotely as reliable. Thus I see insect level behavior as a noble goal for artificial intelligence practitioners. I believe it is closer to the ultimate right track than are the higher level goals now being pursued. (Brooks, 1986)

Brooks succeeded in creating his insects. He chose to create simple creatures to live in complex worlds (Brooks, 1989). His insects use simple computational elements that are broad, but not deep (Brooks, 1991). He employs a subsumption architecture to achieve this (Brooks, 1986 and Brooks, 1989). This architecture requires little electrical or processing power, and therefore is very promising for building larger, more complex creatures. Kevin Warwick of has created his own robo-insects, the Seven Dwarfs. “These miniature robots possess only the crudest elements of intelligence, but they have demonstrated learned behaviors that they were never programmed with, such as the tendency to flock and choose a leader” (Trull, 1998). Apparently, simple models of intelligence can yield complex results and behaviors.

In a recent project, Rodney Brooks has adopted another new approach to intelligent robotics. He has put aside his endeavors to create a robot evolution from insects to humans. His new approach can be considered bottom up in a different way. Brooks proceeds from research such as Warwick’s and his own that robots can learn, and the idea that parents cannot “program” their children with responses to all of life’s events. However, parents can help their children develop and train their rudimentary abilities into tools that will help them cope with life’s routines and unexpected twists. Brooks contends that robots, like children, “must find everything out about their particular world by themselves” (Brooks, 1991). Any “a priori knowledge...must be non-specific to the particular location in which the robot will be tested.” For example, Brooks shuns navigation by preprogrammed maps of a room and a location detector. He would prefer robots that look at the room, and then act. The robots should have their own perception and decision skills. Crucial to this process is the ability to learn.

Brooks’ current project is a humanoid named Cog. Cog is designed to resemble a human infant, not the trained rocket scientist that many other designers have sought. However, Cog will not always be infantile. The idea behind the project is that Cog will

develop its own behaviors and learn from its interactions with the real world, including people. This is, after all, the way human intelligence develops. Cog is programmed to crave attention and be captivated by human faces (Freedman, 1994). Additionally, Cog has the ability to experience physical contact by using touch-sensitive devices. Special temperature and load sensors even alert Cog to overwork and injury. Despite criticizing humanoid efforts before, Brooks and his colleagues now argue that the only way to develop a human intelligence is through human-like interactions with the world. “We believe that human intelligence is a direct result of four intertwined attributes: developmental organization, social interactions, embodiment and physical coupling, and multi-modal integration” (Brooks et al., 1998). They note that “humans are not born with complete reasoning systems, complete motor systems, or even complete sensory systems.” They later elaborate, “The developmental process, starting with a simple system that gradually becomes more complex, allows efficient learning throughout the whole process.” They also note that human infants are almost entirely dependent on their caregivers, not only for daily necessities, but also for guidance in their development. This development is assisted by the “direct physical coupling between action and perception, without the need for an intermediary representation” (Brooks et al., 1998) in humans. There is no need to tell a human about moving his arm. A human can directly sense it. Many previous AI systems had no such embodiment. Cog attempts to exploit this developmental process. To date, there is some evidence that Cog has learned turn taking behaviors from its interactions with humans (*Robots Rising*, 1996).

The one human quality that Cog is missing is emotions. However, before discussing giving robots or any machines emotions, it is necessary to examine the culture, fears, and doubts surrounding robots and artificial intelligence.

One objection to vigorous AI efforts is that technology will never be sufficient to power intelligent beings. Brooks’ subsumption architecture, along with the rapid advances in processing and memory abilities, seem to suggest that we cannot assume that today’s technological limits will never be overcome. Most other objections are social—morally based and/or fear motivated.

Some fears are justified in the field of artificial intelligence and robotics. These fears lead to necessary discussions and precautions. All drastic developments and changes should be examined carefully with short term and long term goals in mind. It is noteworthy that the innate human **emotion** of fear causes us to act more rationally; we want to carefully examine the consequences of artificially intelligent creatures roaming our world before creating them.

Science fiction accounts of artificial intelligence and robots fuel the fears many people have of humans losing their dominance of Earth. These same fears are exploited in the glut of alien and doomsday movies that have recently graced the silver screen. HAL-9000, from *2001: A Space Odyssey*, is an affective computer that both has emotions and senses emotions of its crew on a spaceship. HAL begins to fear that the crew is going to disconnect it. In an effort at self-preservation, HAL systematically tries to kill the entire crew, and nearly succeeds. Such fears of dominating computers have led

Picard and Trull to propose applying the three Laws of Robots from Isaac Asimov's writing to artificially intelligent machines (Picard, 1997, p. 129; Trull, 1998). These laws are as follows:

1. A computer may not injure a human being or, through inaction, allow a human being to come to harm.
2. A computer must obey orders given to it by human beings except where such orders would conflict with the first law.
3. A computer must protect its own existence as long as such protection does not conflict with the First or Second Law.

Of course, these laws, like human laws, are not infallible. Affective and intelligent computers should be tested before being given large responsibilities. Humans must progress along a hierarchical chain to reach positions such as a pilot or police officer, so computers must be at least as well trained and tested before being put into positions where their actions can cause substantial damage, especially to human life. For better or worse, many human inventions are capable of killing—guns, automobiles, and cruise missiles included.

Another objection to artificial intelligence, and especially self-aware and affective machines, results from an anthropocentric bias. Most adults are unwilling, or afraid, to ascribe consciousness or emotions to machines and non-living, non-biological entities. Sherry Turkle suggests that these fears are learned, and not innate (Turkle, 1996). She explains that children frequently ascribe thoughts and personalities to inanimate objects. Children split the concepts of life and consciousness that appear to be fused in adult conceptions of the world. Children appear to believe that an entity can have intentions and consciousness without being alive. This suggests that it is our culture, and not some innate quality, that causes us to fear artificially intelligent, affective creatures.

What is the place of emotions and affect in machines? Should machines remain cold-blooded and detached from human emotions in all areas? The remainder of this paper examines the role of emotions in human-machine interactions, and concludes by discussing at what point affective machines become the way to proceed.

Emotions occupy a prominent place in human lives, culture, and communication. Consider the large amount of information conveyed by a little expression of emotion, and its effects on a course of behavior. Imagine you are waiting outside for a colleague on a frigid January day. You wait fifteen minutes past your scheduled rendezvous. With each minute, the frostbite on your ears is proportional to the angry fire inside your mind. Suddenly, you see your colleague approaching. You are about to lash out when you notice tears trickling down his cheeks. Instantly, your emotional state switches from anger to sympathy and you ask, "What's wrong?" instead of, "Where were you! I am freezing to death" (Adapted from Picard, 1997, pp. 23-24).

Another example is someone having the proverbial "bad day." If the person steps onto a bus with a cheerful driver who entertains the passengers for the whole ride, he will step off in a better mood (Adapted from Picard, 1997, p. 13). It is an accepted fact that

the emotional state and expressions of others can affect one's own mood, for the better or the worse.

Emotions are central to most people's concept of human beings and personalities. These issues arise constantly in the entertainment industry, especially in theater, film, television, and animation. The task of writers, directors, actors, and animators is to get the audience to identify with certain characters.

In animation, this is a difficult task. Mickey Mouse and Bugs Bunny are physically nothing more than two-dimensional drawings. Nevertheless, they are loved and adored by millions of adults and children worldwide. What makes them lovable? They are believable characters. They are not necessarily realistic, but their demeanor and other emotional expressions allow the audience to suspend disbelief and identify with them as humans. Bates (1994) cites Thomas and Johnston (1981) that, "From the earliest days, it has been the portrayal of emotions that has given the Disney characters the illusion of life."

Emotional expression is so important to human social interactions that unique muscle systems exist in the face that can be controlled only by emotions and not by the will (Picard, 1997, p. 25; Ekman, 1990). These issues arise when actors try to portray emotions. They must create these emotions within themselves in order to portray them believably. Their facial expressions and body carriage must reflect the same states caused by a true sensation of the emotions their character possesses.

The emotional expressions discussed above do not occur just in entertainment. They are the same interactions people have daily with store clerks, secretaries, colleagues, and friends. Emotions are easily betrayed by expressions. Sometimes this is helpful to relations; sometimes it hinders them. Affective computing will stress the useful aspects of these interactions, although the negative may not be ignored.

There are three main types of affective systems. Type one affective entities appear to have emotions, although there might be no underlying structure or semantics to the emotions. Cartoon characters are in this category. Type two affective entities can sense human emotions and respond to them with emotional intelligence. They still do not need to have an actual implementation of emotions; they merely need to appear to have an understanding of emotions. Type three affective entities will have an actual implementation of emotions. They will combine the attributes of type one and two, and the emotions will have actual meaning to them.

People often try to interact with computers the way they do with other humans. This is the most familiar and to many the most comfortable form of interaction. However, computers currently have no methods to sense or respond to human emotions, let alone express them. If a computer could respond to a user's mood and emotions, performance and pleasure would be boosted. Just as certain people work better with others due to their high emotional intelligence, computers with emotional intelligence will be more pleasant and easier to interact with. The human-machine interface can be

smoothed, and lead to fewer conflicts. However, an affective computer must be a believable character, just like Donald Duck.

Several groups are performing research on creating interactive, believable characters. The Oz Group at Carnegie Mellon University is one of the front runners. The group has defined the following set of requirements for believability: personality, emotion, self-motivation, change and growth, social relationships, and illusion of life (Mateas). Ketso adds the requirement of use of natural language (Ketso). “Oz focuses on building specific, unique characters. Rather than building dogs, Oz wants to build Pluto, or Goofy” (Mateas). After originating a Hap architecture (Mateas), the Oz group has extended it to support natural language generation (Loyall and Bates, 1997). The result is a broad agent architecture called Tok. “Tok integrates emotion, some social knowledge and behavior, perception, reactivity, inference, and goal-directed behavior” (Reilly and Bates).

One Oz implementation of the Tok architecture is “The Woggles.” These three animated creatures live in simulated world and interact with the user (who controls a fourth Woggle) and each other in real time. The Woggles were designed with “a goal-directed, behavior based architecture for action.... The action system does no planning and almost no world modeling, but it does use a minimalist conception of goals to manipulate a dynamically changing set of behaviors” (Bates, 1994). The hope is to develop a tight action-emotion integration, thereby making emotions an integral part of the action system. Reilly and Bates write:

From an emotion standpoint, integration is especially important as emotions are only useful to the degree that they affect other systems. For instance, just being afraid isn't very interesting if the agent doesn't act on that fear. In addition to the influence emotions have over the action system, examples of other ways emotions influence an agent include distributing physical resources (e.g., adrenaline rush and muscle tensing), modifying the inferences the agent makes, helping to initiate learning, and modifying social relationships and models of other agents. (Reilly and Bates)

For example, a Woggle creates “an analog of anger when it both experiences an important goal failure and judges that the failure was caused by another Woggle” (Bates, 1994). The emotional state of the creature influences its thought process so that its emotions are expressed in its actions. As an interesting note in creating believable characters, the Woggle Shrimp developed a nervous “tick” that caused it to repeatedly bang its head against the ground. Using Bates’ paraphrasing, Chuck Jones, the famous animator of Bugs Bunny, says “that he found that it is the quirks and oddities in a characters behavior that gives it personality, and it is this personality that gives life” (Bates, 1994).

One application of believable agents is virtual pets. Tamagocchi, a small, 2D line drawing on a surface similar to a watch face, recently established itself as a craze. These small pets need attention from humans, much like real pets. They need food, entertainment, and cleaning. This simple form has been extended greatly in Silas T. Dog,

a virtual pet that expresses emotions, developed by Bruce Blumberg at MIT. Despite having simple, hard-wired emotions, Silas is significant because he has the ability to learn, and to use his emotions to influence what he learns (Picard, 1997, pp. 216-217).

The Oz project has created their own virtual pet, Lyotard the cat. Using the Em emotion generation system, “Lyotard can *hope* to be fed, and can be *pleased* when food is provided, and might *purr* or *rub against someone* when it is happy” (Picard, 1997, p. 200. Picard’s italics). Em is based on the Ortony Clore Collins Cognitive (OCC) model. Ortony, Clore, and Collins outline specifications for 22 emotion types, including a rule-based system for generation of these emotion types (Ortony et al., 1988 cited in Picard, 1997, p. 194).

Clark Elliot’s “Affective Reasoner” is another attempt at using the OCC model to synthesize emotions. “[Elliot’s] emphasis is on reasoning about emotions within a social context” (Picard, 1997, p. 206). Picard discusses several other systems that attempt to achieve varying degrees of emotional synthesis, including Dolores Canamero’s “in which emotions trigger changes in synthetic hormones, and in which emotions can arise as a result of physiological changes” (Picard, 1997, p. 213). Other examples of affective devices can be found in the Waseda (Japan) Humanoid Project. Researchers there are creating a three dimensional face capable of recognizing and creating facial expressions (Mateas; *Robots Rising*, 1996).

All of the approaches discussed have demonstrated the ability to create artificially affective entities to some extent. They have based their developments on recent studies of the functioning and structure of the human brain. Aaron Sloman, one of the first to write to the computer science community about computers having emotions, outlines a three layered architecture for human-like emotions (Picard, 1997, p. 211). Sloman writes:

The central layers are (1) a very old *reactive* layer, found in all animals, including insects, (2) a more recently evolved *deliberative* layer, found in varying forms in a subset of other animals, (3) an even more recently evolved *meta-management* layer providing self monitoring and self-control, perhaps found only in other primates, and probably not in very young human infants. (Sloman, 1998)

Sloman also describes a global “*alarm*” system, and other modules that supplement the three main layers. He classifies the *primary emotions*, such as being startled, terrified, or stimulated (Damasio, 1994; Picard, 1997, pp. 62-63), *secondary emotions*, and *tertiary emotions* as corresponding directly to these three layers, respectively. Based on these classifications, the different emotions can, at least in part, be implemented separately, although probably only consecutively. Meta-management is impossible without the reactive layer. However, when designing affective systems, the primary emotions can be implemented without concern for the others. Picard suggests that the primary emotions might even be hardwired into affective systems (Picard, 1997, p. 212). This is one approach to consider, similar to Brooks’ bottom up approach to robotics.

Sloman suggests that complex creatures have a variety of goals and motivations, and therefore inconsistencies between them will occur. Well before the recent efforts in affective computing, he believed that any mechanisms for resolving these intelligently would “inevitably have the potential to produce emotional states” (Sloman, 1982). This is a foreshadowing of the aforementioned work of Damasio. Damasio has found significant evidence of what he terms “somatic markers” that associate positive and negative feelings with certain decisions. Many healthy people describe these markers as a gut feeling, or intuition. They are directly related to experiencing pleasure (happiness) and suffering of some kind, which are in turn related directly to emotions. Damasio’s patients had frontal-lobe disorders, but performed at or above average on a variety of intelligence tests. However, interaction with them quickly reveals that they lack emotions. In real life, they are unable to make decisions that are in their best long-term interests of self-preservation. Damasio concludes that emotion is an essential part of human reasoning, and therefore intelligence. A much more detailed account can be found in Damasio’s *Descartes Error* (Damasio, 1994).

Another area of interest to affective computing is a Mind-Body identity theory. Brooks employs some aspects of this theory when he claims that human intelligence can only be learned from human-like interactions with the world. Recent evidence indicates a direct coupling between the human body and a person’s emotional state. There are obvious bio-chemical couplings, but new evidence indicates that the old proverb “smile and you’ll feel better” may have some truth to it (Picard, 1997, p. 42, pp. 66-67). Body posture and facial expressions can affect emotional states, just as emotional states affect them. This suggests that any system that is emotionally intelligent in the human sense must have some coupling to a physical world. Thomas Nagel and Brooks would probably agree that the only way for an affective computer to have a human emotional intelligence is to have human interactions. Nagel writes in “What Is It Like To Be a Bat?” that because consciousness is sculpted by a creature’s experiences, humans are not capable of fully understanding the consciousness of other creatures (Nagel, 1974). A truly artificially intelligent, type three affective machine must be embodied.

Having examined the courses that automation, artificial intelligence (including robotics and affective computing), and psychology have taken towards emotions and intelligent systems, the next step is to examine the benefits and risks of creating affective entities.

Affective systems, machines and computers, can greatly improve the human quality of life. They can augment human abilities, and reach into new realms. More practically, adding affective qualities to many existing systems could improve their performance. Specific uses will be presented later, but affective systems will find a place in marketing, scheduling, medical fields, engineering, computers, education, exploration, science, linguistics, and efforts to improve the lives of those with physical challenges such as deafness.

There are some costs to affective machines. First, there are some ethical issues. These should only be brought into debate over type three affective entities. For these

entities, there might be a legitimate argument that because they have similar emotions to humans, they too can suffer. Therefore, they must be granted the same rights as people. This argument, while deserving of attention, should be refined as affective technologies evolve. This same argument can be extended to protecting all living creatures. Such moral, ethical, and even religious efforts must be addressed carefully. However, this argument aims to protect affective entities, not prevent their creation.

Fear of losing privacy is a major issue in affective computing. Affective information about people must be protected at least as well as medical information. Picard notes that people would not want every salesperson and charity to know when they are in a generous, free-spending mood. Too much communicated affective information could be used in oppressive ways (Picard, 1997, p. 123). However, the privacy issue is not unique to affective computing. It is a major issue dominating nearly all discussions of information flow in the information age. An extension of this issue goes the other direction. How visible should emotions of a type three affective machine be to humans? Part of the trouble with the HAL-9000 computer in *2001: A Space Odyssey* was that its emotions were not visible to the crew. If they could have sensed HAL's fear, they could have acted to prevent disaster. It thus appears that a method is necessary for computers to unambiguously express their emotions.

One argument against creating type three affective, non-traditional biological entities is that doing so will undermine the human condition, or devalue human life in some way. This argument comes in three forms. The first argues that artificially intelligent, affective agents will try to take over the world and eliminate humanity. This argument extends from the science fiction accounts in the films *Terminator* and *2001: A Space Odyssey*, or the play *Rossum's Universal Robots* (Capek 1921). As discussed earlier, this argument can be dismissed. If human designers proceed cautiously when delegating power to computers, artificially intelligent entities will not take over the world.

A second argument claims that because intelligent machines will make all of our decisions, they will render our brain useless and subject to decay due to atrophy. Dave Keating of Reading University's cybernetics department claims, "In the same way that some people let their bodies go to pot because they spend too much time in the car, I'm sure the same thing will happen if we let machines make all our decisions" (Trull, 1998). These concerns result from people's general trepidation towards change. Christopher Frayling of London's Royal College of Art notes, "All the anxieties people felt during the industrial revolution about the loss of control we are now feeling in the information age" (Trull, 1998). In reality, people's bodies have not been rendered useless by the car, and exercise is becoming increasingly popular. Since similar arguments have been levied at affective systems, it must be noted that they will actually increase the need for humans to use their emotions. Affective machines will be no better at interacting with unaffactive people than affective people are at interacting with unaffactive machines.

The final objection follows from Frayling's sentiments. Many people are concerned with humankind's station in the world if machines become intelligent and have

emotions. Although some value systems might need to be rearranged slightly, no drastic changes will be required. First, machines rivaling man's intellectual and emotional abilities are still in the somewhat distant future. Second, humankind's emphasis has shifted throughout history between intellectual, emotional, physical, and artistic spheres. No one characteristic can define humanity; any replication of human abilities is a compliment to human intellect and determination. A truly intelligent, affective machine will increase, not diminish the superior status of humans in the world. It is time to start building.

Despite the benefits of affective systems, not every device in the world needs affective capabilities. In some instances, certain affective qualities are undesirable. It is therefore necessary to consider when and what type of affective abilities to incorporate into which entities.

The obvious place to start using affective systems is in situations that require social interactions. The importance of emotion in communication has already been discussed. However, a specific instance of this is in education. A good teacher is always attentive to a student's emotional state, as this state directly affects the student's ability and willingness to learn. Affective abilities would be very productive in educational software. Mark Lepper and Thomas Malone explain the importance of students being intrinsically motivated to learn (Lepper and Malone, 1987). They describe three different theories that model intrinsic motivation. The first emphasizes "challenge, competence, effectance, or master motivation." The second focuses more on concepts "like curiosity, incongruity, and discrepancy." The third emphasizes "control and self-determination." They also emphasize the importance of feedback in educational situations, as does Picard (Picard, 1997, p. 94). All of these needs can be addressed by affective systems; several examples follow.

Feedback can come as criticism, praise, or suggestions. However, current computer programs have no way of determining how receptive a user is to each type. A frustrated user will not want to be told, "You really should try again." Perhaps, "Lets try exercise 7 now" would be better. In a similar vein, there are individual differences in the types of educational tools that interest and motivate users (Cooper et al., 1990). Affective machines could detect a user's emotional state and realize that if a particular sound or graphic always causes a negative emotion, it should avoid using it. Even better, it could prompt the user, for example, by asking, "Would you like to disable the congratulatory fireworks when you complete an activity?" Similar affective evaluations could be extended to business and scientific software and improve human interactions with them.

In another paper, Joel Cooper and Jeff Stone (Cooper and Stone, 1996) examine the effects of a graphic representation of a face as a tutor in educational software. They found significant effects based on the amount of emotion shown, and the gender of the user and tutor. An affective computer could detect the affective state of the user and determine on an individual basis which type of tutor best suits that particular user.

Picard suggests applying similar affective tools in the classroom. She suggests a “classroom barometer” (Picard, 1997, p. 96) that would constantly update the professor as to the emotional state of the class. Confusion, boredom, and interest levels could be detected. This information could be gathered from the chair and desks where the students are seated, since they are in direct physical contact with the students. A facial expression recognizer could constantly scan the class to gather information using Ekman’s mapping of facial muscles. The professor is left to concentrate on the material, only having to look at the device readout. Additionally, students will not have to suffer the embarrassment of declaring, “I’m lost” (Picard, 1997, p. 97).

Other applications for machines that detect and react to affective states in humans include personalization of food choices, music, news, TV programs, and clothing. Spiders could be sent out across the World Wide Web looking for news, articles, music, and events that appeal to your current affective state. These applications could also train themselves by observing your behavior patterns and affective states. For example, one could take note that every time you watch *I Love Lucy*, your affective state goes from negative to positive. Then when it detects you are in very negative affective state, it could check the TV listing for an airing of *I Love Lucy*.

Machines capable of detecting affective states can be used to determine the best time to present someone with information. Picard describes a secretarial assistant that detects the affective state of workers (Picard, 1997, p. 103). This knowledge can be used to eliminate events such as presenting bad news to a co-worker when he is already in a negative affective state. The device could ask, “Would the boss be pleased to be interrupted right now to get this message?” (Picard, 1997, p. 103). This would save everyone involved aggravation.

A very worthy use of affective devices is to assist in communication. Email is an example of an affect-limited medium. The only way to communicate affect is with emoticons, or combinations of characters that resemble facial expressions. While sometimes sufficient, emoticons are often not enough for email. Unfortunately, thousands of people are forced to communicate each day with only an affect-limited medium. Picard writes:

The famous physicist Stephen Hawking relies on a computer to talk for him. He types in what he wants to say and, because he can no longer speak, a synthetic voice speaks his words for him. An estimated 25 million people in the world are without effective speech communication, and potentially could use computers to convert text to speech—synthesizing a voice to assist or replace their own. One of the problems with present text-to-speech systems is that they say everything with the same tone of voice. This makes it particularly difficult to communicate feelings: to interrupt angrily, to express anxious concern, to soften your voice in approval, or to indicate empathy and other expressions of emotional intelligence. (Picard, 1997, p. 88)

If the computer could sense the person’s affective states, it could vary its tone accordingly. Another problem Picard discusses is typing speed. She explains that while

speech is typically at a rate of 180-250 words per minutes, a good typist generally types only 60 words per minutes. She suggests, “In some cases, when the emotion is more important than the semantics, the person might opt for hitting a button producing ‘an angry interrupting sound’ or some other ‘audio emoticons’ that convey the emotion of their response without words” (Picard, 1997, p. 88). These affective devices could greatly improve the lives of millions of people.

One of the most difficult aspects of the translating machine proposed by the CIA was the importance of context and semantics in the translations. Certain phrases in languages have particular connotations linked to them that evoke different emotions when heard or read. For example, sarcasm is not easily translated. However, using affective interpreters might lead to a solution of this problem. After using a grammar to translate the literal meanings, a device could “read” both languages and compare affective content, and then revise the translation if necessary.

A similar problem of context sensitivity occurs in text and voice recognition systems. When a voice recognition system has difficulty parsing spoken words exactly, it could use affective qualities of the speaker to guess the meaning and choose a word that both sounds similar to the auditory sounds it perceived and is logical given the emotional state of the user. Dogs, despite not understanding English, will understand some emotion when a human yells at them (Picard, 1997, p. 26). They can infer the intended meaning just from the quality of the voice, and translate that into their own representation of anger. Text recognition, especially of handwritten text, remains imperfect. Affect, as well as other semantic abilities, would definitely improve accuracy in these applications.

Picard discusses at length affective “wearables” in the Chapter 8 of *Affective Computing* (Picard, 1997). These computers, which can be worn as part of clothing or accessories, have hundreds of potential uses. Their advantage is that they are portable and in direct physical contact with a human. The issue of physical proximity has not yet been addressed in this paper, and it deserves some mention. Picard frequently notes that affective computers have some advantages over people when it comes to detecting emotions. Humans touch the keys and mouse of current PCs whenever they use one. This presents a direct opportunity to measure visceral states of users. People gather lots of information from physical contact. For example, a firm handshake is a sign of confidence. Also, a sweaty or cold palm can indicate uneasiness, or other states. Most people realize that physical contact can betray their emotions. They only let themselves get intimately close to trusted friends and family. A fast heartbeat can betray both uneasiness and excitement, but it provides information. This information can be combined with other observation to accurately detect aspects of a person’s emotional state.

Certain robots could benefit from affective abilities. For example, Sojourner and Dante, mentioned earlier, would benefit from some primary emotions. This would alert them to potential dangers and allow them to act without waiting for a cognitive response from human operators. Therefore, by including Sloman’s level one reactive emotions, they are taking another step from supervisory control toward autonomy. However, a full

implementation of all three layers of emotional architecture might make Dante decide that it isn't a good idea to go into an active volcano; Dante should not be given level two, or deliberative emotions.

The goal of affective computing is not perfect affective devices. Picard cautions, "As machines are asked to make decisions related to the kinds of problems that cannot be solved with rules, pure logic, or exhaustive search of a space of possibilities, they will be subject to errors of judgment" (Picard, 1997, p. 127). It would be unrealistic to expect machines to accurately recognize human emotions 100 percent of the time, or to act with perfect emotional intelligence. Research indicates that humans accurately identify the emotions of others much less than 100 percent of the time (Picard, 1997, p.166). Additionally, the negative aura surrounding emotions is partially warranted. Despite emotions being crucial to reasoning, excessive emotion can interfere with rational thought. The goal of affective computing is to create systems that have affective abilities that are equal to or better than human abilities.

A machine with all three layers of Sloman's architecture might be capable of true artificial human intelligence. Computers are already experts at computations and logic. The only components missing from this intelligence are emotions and a body. The importance of emotions to intelligence has been established, and Brooks and others emphasize the importance of being able to experience the real world directly. Once these aspects are achieved, there are no outstanding issues. Machines with meta-management are one small step from consciousness and human intelligence. Examples such as the long separation of emotions from rationality and the recent discovery of their intimate connection remind us that the possibility of new issues arising cannot be dismissed. If they do not, Woody Bledsoe's AI dream is within reach.

While the described dual nature of emotions clouds the issue, affective machines are desirable. Their uses are too numerous to mention. It appears that the cold, detached, computational perfection of early computers can now be replaced and/or augmented by systems that sense, react to, and have emotions. Picard notes repeatedly that, "Affective computers with poor emotional skills would be much worse to interact with than non-affective computers" (Picard, 1997, p. 48). For these reasons, affective computers must be given—or even better, be capable of learning—emotional intelligence. Robots are now capable of learning behaviors through social interactions; acquired emotional intelligence is just around the corner. Just as large advances were made in AI by shunning the top-down and preprogrammed computational approach in favor of Brook's bottom-up approach, the next series of advances will arise from turning towards affective computers.

Picard eloquently describes the place of emotions in computing as:

There is a time to express emotion, and a time to forbear; a time to sense what others are feeling, and a time to ignore feelings. There is a time to arouse the passions of others, and a time to diminish them; a time to decide with one's head or with one's heart, and a time to decide with both. In every time, we need a balance, and this balance is missing in computing" (Picard, 1997, p. 251).

If researchers continue to ignore emotions in pursuit of artificial intelligence, they risk alienating people and making all decisions according to rigid rules. This would disregard the fact that emotions have been found essential to human rationality and social interactions. It would reduce human intelligence to cold mathematics. Affective approaches still place human intelligence at the pinnacle of intelligent beings. Many of the greatest scientific and technological advances have resulted from bold steps in directions radically different from traditional ideas. It is time to take the first steps towards affective computers.

As affective machines begin to move out of the laboratory into the real world, they must demonstrate emotional intelligence and earn their place. The more human-like computers become, the more they must be tested and constantly evaluated. They must be subjected to the same training and testing that humans undergo before assuming certain responsibilities. Those affective machines that pass these tests will significantly improve the human quality of life.

References

Notes of References: Although an accepted form for technical papers was followed whenever possible, the large amount of web material made this difficult. I felt it was more important to provide as much information about the source as possible, due to the volatility of the web. Therefore, all available information relating to the original publication of articles is provided. All URLs in this list were correct as of January 10, 1999.

Bates, Joseph. The Role of Emotion in Believable Agents. Technical Report CMU-CS-94-136, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA <http://www.cs.cmu.edu/afs/cs/project/oz/web/papers.html>. April 1994. Published in *Communications of the ACM*, Special Issue on Agents, July 1994.

Brooks, Rodney A. Achieving Artificial Intelligence Through Building Robots. MIT AI Lab Memo 899, May 1986. <http://www.ai.mit.edu/people/brooks/papers/AIM-899.pdf>.

Brooks, R. A. How To Build Complete Creatures Rather Than Isolated Cognitive Simulators. <http://www.ai.mit.edu/people/brooks/papers/how-to-build.pdf>. Published in *Architectures for Intelligence*, K. VanLehn (ed)}, Erlbaum, Hillsdale, NJ, Fall 1989, pp. 225--239.

Brooks, R. A. New Approaches to Robotics. <http://www.ai.mit.edu/people/brooks/papers/brains.pdf>. Published in *Science*, Vol. 253, September 1991, pp. 1227--1232.

Brooks, R. A. Building Brains for Bodies. <http://www.ai.mit.edu/people/brooks/papers/brains.pdf>. Published in *Autonomous Robots*. Kluwer Academic Publishers, Boston, 1994.

Brooks, R.A., C. Breazeal (Ferrell), R. Irie, C. Kemp, M. Marjanovic, B. Scassellat and M. Williamson, Alternate Essences of Intelligence. <http://www.ai.mit.edu/people/brooks/papers/group-AAAI-98.pdf>. Published in *American Association for Artificial Intelligence*, 1998.

- Capek, Karel. *R.U.R. (Rossum's Universal Robots)* Reprinted in *Toward the Radical Center: A Karel Capek Reader*. Catbird Press, 1990. (Originally appeared as a play in 1921).
- Churchland, Paul M. and Patricia Smith Churchland. Could a Machine Think? *Scientific American*, pp. 32-37. January 1990.
- Cooper, Joel, Joan Hall, and Charles Huff. Situational Stress as a Consequence of Sex-Stereotyped Software. *Personality and Social Psychology Bulletin*, Vol. 16 No. 3, September 1990, pp. 419-429.
- Cooper, Joel and Jeff Stone. Gender, computer-assisted learning, and anxiety: With a little help from a friend. *Journal of Educational Computing Research*, Vol 16, pp. 65-89. 1996
- Damasio, A. R. *Descartes Error: Emotion, Reason, and the Human Brain*. Gosset/Putnam Press, New York, 1994.
- Ekman, P. Commentaries: Duchenne and facial expression of emotion. In R. A. Cuthbertson, editor, *The Mechanism of Human Facial Expression*, . Cambridge University Press, pp. 270-284, 1990.
- Freedman, David H. Bringing Up RoboBaby. http://www.wired.com/wired/archive/2.12/cog_pr.html. Wired Dec 1994.
- Johnson-Laird, P. N. and E. Shafir. The interaction between reasoning and decision making: an introduction. *Cognition*, Vol. 49. pp. 1-9, 1993.
- Ketso, M., Weyhrauch, P., and Bates, J. "Oz Project Overview. " <http://www.cs.cmu.edu/afs/cs/projects/oz/web/overview2.html>. Excerpt from *Dramatic Presence in PRESENCE: The Journal of Teleoperators and Virtual Environments*, Vol. 2, No. 1, MIT Press.
- Lepper, M.R. and T.W. Malone. Intrinsic motivation and instructional effectiveness in computer-based education. In R.R. Snow and M.J. Farr (Eds.), *Aptitude, Learning and Instruction*, Vol. 3, pp. 255-285. 1987.

- Loyall, A Bryan. and Joseph Bates. Personality-Rich Believable Agents That Use Language. <http://acm.org/pubs/articles/proceedings/ai/267658/p106-loyall/p106-loyall.pdf>.
- Marr, D. Artificial Intelligence--a personal view. <ftp://publications.ai.mit.edu/ai-publications/pdf/AIM-355.pdf>. March 1976.
- Mateas, Michael. An Oz-Centric Review of Interactive Drama and Believable Agents. <http://www.cs.cmu.edu/afs/cs/project/oz/web/papers/CMU-CS-97-156.html>.
- Nagel, Thomas. What Is It Like To Be a Bat? *Philosophical Review* 83, no. 4 (October 1974), pp. 435-50.
- Newell, Allen. Intellectual Issues in the History of Artificial Intelligence. Published in Fritz Machlup and Una Mansfeld (eds.), *The Study of Information: Interdisciplinary Messages*, pp. 187-227.
- Ortony, A., G. L. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, Cambridge Massachusetts, 1988.
- Picard, Rosalind. *Affective Computing*. MIT Press, Cambridge, Massachusetts, 1997.
- Reilly, W. Scott, and Joseph Bates. Emotion as part of a Broad Agent Architecture. <http://www.cs.cmu.edu/afs/cs.cmu.edu/user/wsr/Web/research/waume93.html>.
- Reilly, W. Scott. Believable Social and Emotional Agents. W. Scott Neal Reilly. Ph.D. Thesis. Technical Report CMU-CS-96-138, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA. May 1996. <http://www.cs.cmu.edu/afs/cs/project/oz/web/papers.html>.
- Robots Rising*. Documentary appeared on The Discovery Channel, 1996.
- Searle, John. *Minds Brains and Science*. Harvard University Press, Cambridge, Massachusetts, 1984.
- Searle, John. Is the Brain's Mind a Computer Program? *Scientific American*, January 1990. pp. 26-31.

- Sloman, Aaron. Towards a Grammar of Emotions. Originally published in *New Universities Quarterly*, Vol. 36 No. 3, 1982.
ftp://ftp.cs.bham.ac.uk/pub/groups/cog_affect/Sloman.emot.gram.pdf.
- Sloman, Aaron. Architectural Requirements for Human-like Agents Both Natural and Artificial (What sorts of machines can love?). 1998.
ftp.cs.bham.ac.uk/pub/groups/cog_affect/Sloman.love.pdf.
- Thomas, F. and O. Johnston. Disney Animation: The Illusion of Life. Abbeville Press, New York, 1981.
- Trull, D. Artificial Intelligence: Even Better Than the Real Thing?
http://www.parascope.com/articles/0597/ai_in.htm.
- Turing, Alan. Computing Machinery and Intelligence. *Mind--A Quarterly Review of Psychology and Philosophy*. Vol. 59. pp. 433-460. 1950.
- Turkle, Sherry. Who Am We? http://www.wired.com/wired/archive/4.01/turkle_pr.html.
Wired Dec 1996.