Getting Emotional:

A short survey of modern thought on machine affective states

Kyle Wheeler

May 2, 2003

Contents

1 Introduction

When talking about computers and emotion, there are two primary questions. The first of those questions is why would we want computers to be able to understand or express emotions, and the second question is how can computers be made to understand or express emotions. These two questions are addressed by the writings of Dr. Aaron Sloman and Dr. Rosalind W. Picard.

2 Aaron Sloman

Sloman's primary contribution to the pursuit of general intelligence and affective computers is his architectural model of how a mind could work—a model which he believes to be superior to the ideas of everyone else in the field. He addresses some of the basic philosophical issues regarding machine intelligence, with a strong conviction that intellect and intelligent systems are pure information processing machines needing only the proper architecture to achieve intelligence—an intelligence which can operate independently of emotion, but may use emotion in the absence of sufficient processing power as a shortcut or as an extra input.

2.1 Architecture-Based Conception of Mind

2.1.1 The Architecture

In this paper, Sloman introduces his model for intelligent information processing. His model combines the three-column approach to intelligence [3, p3]—in which information is taken into a system by sensors, evaluated by processors, and then output of some kind is generated and expressed via effectors—with a three-layer approach [3, p3] to intelligence—which features a reactive layer at the bottom to handle quick things like reflexes and "muscle-memory"-style activities, a deliberative layer above that to combine reactive-layer types together and to tweak reactive-layer control points in order to perform more complicated actions that can be described as a slightly higher form of abstraction or grouping of smaller steps that compose a more sophisticated action, and a meta-deliberative layer which performs the same actions

as the deliberative layer does, but upon the deliberative layer. This can more easily be described as a three by three grid, with sensors along the left side, effectors along the right, processors down the middle, reactive along the bottom, meta-deliberative along the top, and deliberative through the middle.

This architecture alone is worth considering as interesting, but Sloman adds another piece onto it (in a sense, onto the front of it) which is a pervasive alarm system where any one location in the grid may send a priority communication to any other location on the grid [3, p4]. This, ostensibly, handles reflexes in a more useful and flexible manner, as reflexes can be both physical (in other words, purely reactive), or may be logical or emotional (for example, a sudden surprising thought may trigger jumping out of your chair).

2.1.2 Philosophical Musings

In addition to introducing this architecture, Sloman spends a great deal of time discussing how confused most discourse on the subject of mental architectures and emotions has become both because of how extensively the area relies upon ill-defined cluster concepts [3, p2–4] and how new the concept of information processing is relative to the concept of matter and energy processing [3, p7]. He proposes more useful ways of thinking about cluster concepts like intelligence, emotion, pleasure, and understanding, among others. One way Sloman suggests is to stop considering things as continuous gradients between two extremes, and instead as discontinuous, segmented regions between two extremes. For example, evolution must be discrete, he explains, because it is implemented by a finite number of genes which cannot vary continuously. The same is true of learning, he argues, as learning is a conglomeration of discrete steps like the formation of processing modules, the creation of links between modules, the transfer of control of processing between modules, and so forth [3, p13-14].

Sloman also makes an interesting claim regarding the causality between the mental and physical entities, drawing a correlation to virtual machine systems [3, p10-11]. The critical analogy he makes is that the mental body (i.e. the mind) affects the low-level in the same way that a software program affects the bits and voltages of a computer. Also, a software program may contain a virtual machine, and thus may "run" other software programs by emulating or simulating what might actually happen while at the same time making either something similar or something very different happen in the actual hardware. Thus, just as a software package has logical cause and effect that is independent of the actual voltages of the hardware, "causal completeness at a physical level does not rule out other kinds of causation being real," according to Sloman, essentially equating the mental entity with a simple logical abstraction that has it's own causes and effects which are reflected in the lower-level physical body only because that is how it is implemented. Because the physical body is a full implementation, it has it's own full-causation, but this is simply the implementation of larger logical ideas, structures, and processes (in other words, architecture) embodied in the mental entity.

2.1.3 Criticism

Many of the concepts introduced in this paper are far-reaching, as one would expect from a paper outlining an architecture rather than an implementation. Nevertheless, Sloman expresses perfect confidence that the current problems faced by artificial intelligence researchers in their quest for human emulation will be surmounted. It is perhaps this confidence that is one of the primary criticisms one can make of the paper. For example, he supposes that robots will one day wonder whether humans are conscious [3, p10]. This prompts the reader to wonder, however, how wondering fits in Sloman's proposed architecture, and the answer is not obvious. The meta-management layer, as explained, seems rather self-directed, rather than externally directed.

Other criticism of the paper can be directed at Sloman's philosophical musings. Sloman attempts to put to rest the common "Zombie" problem, that is, the concept of entities that may have all the necessary and relevant functional units required for intelligence and perception (much as humans have), but that still don't experience things, as qualia. His argument, however is simply to assign the definition of intelligence and the definition of being able to experience qualia to the state of having the architecture for experiencing it, which is essentially a semantic argument, rather than a logical one [3, p11]. This, he admits, will not convince anyone who does not already agree with his assertion that humans are no more than information processing systems, and makes no further arguments supporting that point.[1] While

[1] Sloman labels conflicting opinions "incomprehensible"

Sloman explains thoroughly that humans process information as a matter of course, he does not successfully convince the reader that humans are nothing more than information processors, although he relies on this assumption [3, p7–8]. Similarly, in his conclusion, Sloman points to the example of chess-playing programs and theorem-provers, stating that they neither enjoy nor get bored by doing what they do because they lack the architectural structure for enjoyment or boredom, even though they may be programmed to give the appearance of such emotions. However, he does not take the opportunity to further argue that programming an architecture for boredom is anything more than a roundabout way of programming to give the appearance of such emotions.

2.2 Beyond Shallow Models of Emotion

This paper, as is declared on the very first page is not intended to be scholarly, but rather to provocatively establish Sloman's opinion on the issues inherent in emotion generation and simulation. It also serves as a reasonable conceptual summary of many other papers that Sloman has written. Much of the paper is a re-description and re-justification of the three-layered three-column approach described in the previous paper.

2.2.1 The Basis for the Architecture

Sloman begins with describing shallow models of emotion, how they are currently used, and in what forms he finds them acceptable. Essentially, he relegates them to toy status, useful for educating or entertaining and little else [4, p2–3]. At the same time, Sloman describes the general semantic problems that are generally raised in nearly any rational discussion regarding emotions, namely that many of the critical terms (such as "emotions" themselves) are poorly defined [4, p3–4] (if at all). Such confusion in even the definition of something as basic as what exactly an emotion is raises the obvious question of what to do with emotions and whether they are or can be useful at all without a concrete definition. Sloman proposes that there are four possible ways of dealing with the problem of this ill-defined terminology: first, ignore emotions; second, invent a precise definition of emotion that can be easily dealt with and implemented; third, treat emotions as probabilistic attractors that influence decision making with probability functions that can be altered as a part of learning or to model observed behavior in humans; and finally, invent a deep theory of information processing that exhibits behavior similar to emotions or uses states and processes and state transitions that can be labeled as similar to emotions (an approach which sounds somewhat similar to the second approach). The first approach is the direction that has been taken already, and is generally uninteresting for the emotionally oriented researcher. The second is too simplistic, and ends up (at least in theory, trivializing the full breadth and depth of what emotions really are. The third also seems to trivialize emotions, and does not serve as a good model for basing research into our own psyches—which is a partial goal for emotional models. The last approach is therefore, quite obviously, the route Sloman wishes to explore [4, p4].

In exploring this method of handling emotions, Sloman starts by explaining some of the shallow models that have already been theorized and explored—in particular, the two shallow models he believes can be extended. He explains the three-tower and three-column approaches in terms of their relation to our conceptions of how the human mind works (as this is the ultimate goal of general artificial intelligence), and explains why they are probably the most useful of the shallow models. For example, the three-layered approach makes a lot of sense when compared with and presumed to be directed by evolution, with simple reactive systems like flies embodying the most basic layer, and more complicated animals that appear to be capable of logic and goal-directed activity like dogs embodying a system with both a reactive and a deliberative layer [4, p11]. Humans, of course, are supposed to embody the full three-layers. As well, the three-column approach is a logical way of thinking about things that doesn't constrict the implementation much—that is, a separation, conceptually, of thinking, feeling, and acting. The model puts the sensors first, feeding input into the processing or thinking part, which then sends output to the effectors [4, p9–10]. One could say this architecture is the embodiment of "think before you act," and has a basic structure that agrees with the notion that intelligence and human brains are essentially data-processing machines. With these two architectures in hand, Sloman combines them, making a nine-piece grid. In attempt to more closely model how he believes the mind to actually work (modeling mechanisms like the spinal cord and the limbic system), he adds a system of alarms, allowing any part of the

[3, p11].

grid to, in computer architecture terms, send an interrupt to any other part of the grid [4, p12-13]. This system, he argues can scale with evolution, all the way down to something with just a simple reactive layer, like a fly (although the utility of a system of alarms in a purely reactive system seems redundant) [4, p14]. Sloman also suggests simple extensions to this architecture to allow it to model all sorts of different intellectual structures, such as threshold-based attention filters in between grid blocks (protecting the deliberative layers from lower, higher frequency interrupts, for example), memory structures, and motive generation (an idea that may or may not be useful to make separate from the rest of the deliberative systems) [4, p15].

2.2.2 Emotion's Influence

The final subject the paper addresses is whether or not emotion is actually required for an intelligent system to be intelligent. It has been suggested by several researchers that perhaps any system that is sufficiently complicated to be considered intelligent will necessarily generate states of existence that are analogous to or can be considered to be emotions. Sloman is directly opposed to this notion on basic logical grounds, claiming that emotions and intelligence are logically separate and one does not require the other. He gives a useful example: that of a computer. In many operating systems with modern memory management schemes, there is a state they can get into called "thrashing" where most of the computer's efforts are put into swapping memory pages between main memory and external storage. However, the fact that the system can get into this state (if sufficient memory is unavailable or too many things are attempted at once) is not a requirement for a useful operating system, and likewise, Sloman argues, it is not a requirement for a system to be intelligent that a system have emotions [4, p17].

2.2.3 Criticism

Sloman may be correct on this point, philosophically and logically, it is possible to imagine a useful computer system that is incapable of "thrashing" (the computer system aboard the Starship Enterprise, for example, does not seem to have this problem). Similarly, it is possible to imagine intelligent beings without emotions (for example, Vulcans). It is also possible that such imaginings about logical possibilities are as useful and as valid as imagining creatures that have all the necessary architectural structures for intelligence and emotion and experience and yet do not actually experience anything [3, p11]. Both are conceivable, given a sufficiently powerful imagination; and yet both may be perhaps just that—purely imaginary and in reality perfectly ridiculous. Perhaps, in order to be useful, a computer must have a modern memory system capable of relocating blocks between fast memory and slow memory, and thus be capable of thrashing, as a matter of course. Similarly, perhaps the ability to die is a requirement of life. Not because it is impossible to think of life without death, but because for all practical purposes, it is impossible to create life without the possibility of death.

3 Rosalind Picard

Picard does a good job summing up the challenges and justifications for artificial emotions in intelligent (and even unintelligent) systems. Questions she tackles rather thoroughly (and that seem to be some of the most interesting) are ones like what it means for a computer to have or understand emotions, why would such a thing be useful or even desirable, and what problems may arise if we actually succeed.

3.1 What does it mean for a computer to "have" emotions?

The topic of this paper is fairly self-evident from the title. Picard makes it clear that in her opinion, the main purpose of having a computer understand or "have" emotions is to make it less frustrating to interact with. The ultimate goal, for her is to have a computer that can detect when the user is frustrated and can respond appropriately to make the user less frustrated. That said, Picard breaks emotion down into four things, or capabilities, that a human has as part of emotions, and discusses what each means in terms of a computer's ability to replicate it. These four capabilities are emotional appearance, multi-level emotion generation, emotional experience, and mind-body interactions [2, p2].

3.1.1 Emotional Appearance

The first component, emotional appearance, is perhaps the easiest of the four for computers to replicate—to a certain extent, they already do in

many cases. Programs for little children, for example, frequently use pictures and sounds of encouragement to express praise and satisfaction with the child's progress. On the other hand, because it is so easy to do, it is tricky to say that it necessarily means anything [2, p4]. Much like a person, a computer can put on a happy face without actually being happy. The computer lacks real involuntary emotional expression—all emotional expression is something that has been carefully calculated, rather than a state of the machine affecting how it behaves in an unconscious manner. And perhaps the issue boils down to computers being incapable of a subconscious (at least at the moment). In any case, it is certainly true that emotional expression is easily faked, and means very little about whether or not any particular emotion (or emotion at all) is really happening behind the expression.

3.1.2 Emotion Generation

This lack of motivation problem with emotional expression leads directly into the next part. In order to be able to say that the computer is actually expressing an emotion that it can minimally be able to lay claim to possessing, it must have some sort of method for generating emotion, in other words, it must have an emotion selector. The basic modeling technique that Picard brings up is a dual-speed system [2, p5]. Essentially, emotions seem to be generated at two general speeds, depending on the situation—a quick and dirty (and possibly inaccurate in some situations) method for critical and self-preservation-related situations (fear), and a slower (and generally more accurate) method for more thoroughly reasoned—though not necessarily consciously reasoned—emotions. This is not to say that these two categories cannot be subdivided further, or that these are the only ways that emotions can be generated, but merely that emotional speeds seem to cluster in a bipolar manner. Neuroscientists may discover more accurately how emotions are generated, which may lead to more accurate models, but as a basic sort of division, this simple speed delineation is fine.

3.1.3 Emotional Experience

A system with simply an emotion selector and a method for expressing the selected emotion, however, is lacking something more philosophically fundamental, and that is the emotional experience, or expressing the emotion to oneself (not consciously, of course) in addition to others. Of course, there is an obvious distinction between our own human experience of emotion and a computer's experience of emotion, and they cannot be the same (much as the fundamental difference between the experience of a bat and the experience of a human), but regardless, Picard sees no method for giving computers this ability as yet [2, p7]. While we can easily foresee implementing many of the functions of consciousness and attaching self-monitoring sensors, it is unclear whether this will create an actual experience to existence or not. Picard recognizes the larger problem of talking about emotion and intelligence and whether or not we can imbue machines with these qualities because they are ill-defined words, and the concepts are poorly understood even when discussing them in relation to ourselves (some philosophers remain concerned that equally as much as we cannot demonstrate that robots could possibly have such qualities, we cannot similarly demonstrate that we ourselves can possibly have them—hopefully, if we figure out one half of this problem, we will have the other half). Picard also remarks very briefly on the inherent linguistic problems with saying that the computers are being "given" emotion, or can "understand" emotion, when they may only be imitating it.

3.1.4 Mind-body Interactions

One of the more interesting aspects of systems that have emotions (that we're aware of and can monitor) is that emotions frequently affect and are affected by other bodily activities. People feel happier when they smile, and sometimes shake uncontrollably when afraid or angry. More than that, changes happen at the chemical level in our bodies, causing or caused by emotional changes (adrenaline or Valium, for example). Computers would have to be very complicated indeed to replicate this kind of mental-physical interaction. Beyond even that, though, Picard notes that emotions can selectively modify behavior to a large degree (her example is that of loving and lying, where behavior does not change much when expressing love whether lying or telling the truth but does change significantly when expressing anger, based on whether lying or telling the truth [2, p8]). Pain is also a good example of the interaction between mind and body and how "real" it must be—Picard brings up the example of people who have lost the ability to feel pain and so have pain sensors that communicate pain through the use of annoying sounds. The problem with such artifi-

cial pain systems, however, is that people tend to ignore them or turn them off, indicating that they are not real enough to get the necessary attention [2, p8–9]. Thus, for a computer, if it is to have a similar system, the "pain" (or whatever emotion) should not be able to be ignored or turned off except under extreme circumstances (i.e. greater goals or at the behest of the machine's designer).

3.1.5 Philosophical Musings

Perhaps the largest problem, or the key problem, to affective machines, is that of the emotional experience. While we like to think that by giving a computer "emotions" we are giving them some sort of objective thing, or ability, it may be that "natural" emotions for a computer are entirely different in content, to the point where they may not be resemble human emotions at all. Thus, attempting to give computers human-style emotions may be like trying to feel the emotion of a dolphin, and thus may be a futile gesture at the ultimate extension of anthropomorphism. On the other hand, we can give computers every (known) appearance of human emotion that we can, and still not be certain, philosophically that they have emotions. This is similar to our own inability to prove that other people have emotions, however we simply assume that they do because they are so similar to us in physical structure. With robots and other machines we cannot make this assumption, and we are left with our vague behavioral definition of emotion to try and quantify whether emotion exists in another entity.

3.1.6 Criticism

Picard couches her arguments in very conciliatory and general language, and in so doing avoids nit-picking criticism, but at the same time avoids saying much that is particularly new or useful. Where she addresses the main philosophical problem with machine emotion—the experience of that emotion—Picard bunts, leaving the problem up to the inscrutable external entity of the soul [2, p10]. She does not address the more basic assumption that is being made in assigning emotions to even other human beings, simply taking for granted that other humans feel emotion and that the fact that they do this is so obvious that it is beyond question. She would do well to address the concerns of B. F. Skinner, who said, "The real problem is not whether machines think, but whether men do."

Picard's motivation for pursuing emotionally-aware computers is, ostensibly, to reduce people's frustration with computers [2, p2]. However, frustration with computers generally stems from a misunderstanding between the user and the machine—where the user makes an assumption, the machine does not, for example, and when examined carefully, the user (or the original programmer) was really telling the machine to do exactly what it did. At the very least, adding emotion awareness to a computer will not fix this problem, and may exacerbate it, as emotion is now another axis along which misunderstanding can occur.

3.2 Affective Computing

The books is divided into two main sections. The first discusses the philosophy of, the motivation for, the applications for, and the concerns about: affective computing. The second half of the book is much drier than the first, and talks about how computers can be built to recognize and synthesize emotions, and how we can use them.

The first section of the book raises some very interesting points about how people express emotion and how those emotional cues could be better used to our own benefit.

3.2.1 Ranged vs. Pervasive Sensing

One of the first things that Picard cheerfully puts to rest is the notion that computers should detect emotion the way that we do. It is a common conception that in a general-purpose emotion detector the only required input should be the same as our own built-in general-purpose emotion detector, which is to say, video, audio, some smell, and in a rare circumstance, tactile—generally all from a distance. Such goals are rather lofty, and Picard is convincing that as computers and sensors get smaller and more ubiquitous, there is no valid reason to *not* use more advanced sensory techniques for more accurately gauging people's emotional state. Plus, emotional cues can be very subtle—as subtle as a small timing change in the speed of a person's gait, a minuscule extra hand gesture to close a door quickly, or even simply a glint in a person's eye (subtleties which can completely invert perception of the true emotional state) [1, p28,30]. Computers need, if we're going to be practical about it in the near future, all the help and detailed information they can get. On top of that, in theory, the more information about a person's physical state the system has, the more accurate it can be in its estimation of that person's emotional state. As humans where nearly all

of our emotional detection senses ony function from a distance, we make mistakes. Frequently we misinterpret someone's actions as meaning something other than as they were intended. Perhaps if we had more sensory information about the state of the person in question, we would be more accurate—and the same holds true for computers.

3.2.2 Context

It is an open question whether the additional information is a direct result of an emotional state, or is merely extra subtle context that the person's emotions work in (is the person sweaty because they're worked up about something, or is the sweat irritating them), although the effect is essentially the same. This does bring up the point Picard makes about the importance of context in determining the emotional state of an individual. Context can range from information like the age and gender to culture (or inhibition) of the person in question and their history of behavioral–emotional mapping. The history of behavioral–emotional mapping is perhaps the most important, as one person may twitch when nervous while another may only twitch when extremely angry, and a third person may twitch all the time without thinking about it. The individual differences are enormous, and may not relate very well across people. Picard suggests that the solution, at least for the short term, may be to copy the techniques of voice-recognition programmers, that is, to focus on finding the correct behavior-to-emotion map for a single individual. [1, p32–34]

3.2.3 Emotions and Cognition

Picard notes that cognition can greatly affect emotions, and as such must be taken into account for any system which models emotions. Her example is that of a person who is hot and irritable and gets hit in the back of the legs really hard. Upon turning around, feeling very mad because of the physical attack, the person discovers it was a woman in a wheelchair who lost control. Probably, the person in question will not be mad the instant he realizes that the physical strike was in no way intentional and that the woman in the wheelchair probably needs more attention. With this example, Picard introduces the idea of primary and secondary emotions, or tier one and tier two emotions. Tier one (primary) are knee-jerk reflex "emotions" such as fear, startle, quick anger, and so forth. Tier two emotions are ones that come only with a little bit of thought, such as grief, slow anger, sorrow, and contentment. [1, p35–36,62]

The idea of the two tiers is an interesting one, and prompts one to attempt to figure out the utility of such a system. Tier one are emotions that seem to happen in humans as well as lower life forms (like dogs and cats), while tier two are a little more developed in humans than in other forms of life. It is interesting to note that tier one emotions seem to all be types of cognitive shortcuts (happens without cognitive intervention, as a result of physical somatic responses), to help with survival instincts. Tier two, however, seems more tied to learning than anything else (Picard has an example of a man who could not experience tier two emotions, and as a result could not learn from his mistakes [1, p37]). Of course, this learning introduces the deadening effect (if a person is exposed to a given stimulus too much, the response gradually shrinks in intensity) and the possibility for emotional detachment—something which doesn't seem so possible with the primary emotions. For example, when giving a scholarly lecture, some emotions can be suppressed (typically the tier two emotions) and some cannot (tier one).

Picard does make reference to the apparent utility of emotions in boosting creativity. She also brings up the interesting characteristics of willful versus inherent emotional expression—namely that they take different paths through the brain, as evidenced by brain damaged patients who can smile at a joke but not on command, and vice versa [1, p41–42]. One interesting conclusion Picard didn't specifically state but that seems obvious is that sympathy seems to stem from the way events are remembered, namely that events are associated with the primary emotion experienced at the time which makes good memories easier to remember when in a good mood and bad memories easier to remember when in a bad mood [1, p41]. Thus, when a friend is in trouble, it is easier to remember times when we were ourselves in trouble than it is under happier conditions.

3.2.4 How Would Emotional Computers Be Useful?

It seems that the obvious instances of computers having emotion would not be useful—for example, HAL from 2001 being unable to converse and understand people's emotional responses in any more than a completely naive way, or a computer getting irritated with you when you input the same wrong thing several times is obviously

not desireable. However, Picard makes the case (several times and in several ways) for the usefulness of computers that can recognize and/or produce emotional responses.

The benefit for emotional recognition is clear. From recognizing anger in cars (and playing soothing music) to recognizing frustration at the terminal (and suggesting a break) to judging when a good time for an interruption would be [1, p103] to helping autistic people recognize and express emotion [1, p89], the benefits are everywhere. Perhaps some of them, like changing the music to calm down the driver of a car seems a bit manipulative, but there's no denying the usefulness. Picard talks briefly about affective transmission in communication protocols (like email, instant messenger, and the telephone) [1, p87] and how emotional recognition might possibly be used as a compression technique for pictures of faces or of other things, but her arguments are not particularly convincing; in any medium other than video, we generally prefer to have control over the emotion that is expressed, and such emotional compression hardly seems much of a benefit (if it is even practical), being simply tacked onto existing modes of communication.

The benefits a computer might get from emotions and emotional recognition can be broken into two categories: internal benefits and external benefits.

Using emotions internally, of course, does not require that a computer express these emotions externally. Emotions can still be useful to the computer for many of the same reasons they are useful to humans. For example, emotions are sometimes used as shortcuts to processing, either for speed reasons or for bulk-processing reasons. The fear response may cause the computer to behave differently to save itself without specific reasons and without careful thought (behaving differently to, say, avoid a viral infection or a steep drop). Emotions, good and bad, can also be used to prioritize large bulks of information quickly, so as to deal with and/or avoid information overload. Picard also points out the relation between emotion and memory, which may be an integral part of intelligence. Also, while it may not be a strict benefit, but a useful ability for the computer to have if it has emotions is the ability to be aware of its own emotions and to manage them [1, p77]. If a computer can be aware of its own emotions, it can reason about them and can even use them as input to decision making (or as motivation for doing or not doing things).

Using emotions externally is a different issue. In some way, as indicated by Picard, computers may be better at expressing emotion externally (through the use of pictures and cariacatures) than humans are, even to the point of being able to express contagious emotion (happiness or depression, for example).The real question is what kind of spontaneous emotion can be displayed, since the most obvious emotional indicators that a computer could employ seem far too intentional and easily (and perhaps intentionally) misleading. An interesting example of a robot spontaneously indicating it's internal state (similar to an emotion) that Picard mentions is that of a robot demonstration where at some point during the demonstration the robot simply stopped. As it turned out, it stopped because its input buffers were full, which could be viewd as a particularly robotic affective state, and stopping was a spontaneous expression of that. Humans allow emotional influence to go the other direction as well—smiling can make humans feel happy, for example. Computers can have sensors to attempt to simulate that direction of emotional influence, but there seems to be something far too contemplative about that.

3.2.5 Possible Problems

As with all things, where there are benefits, there are also liabilities, and any system involving computers with emotion is bound to have at least some down sides. Picard brings up the major problems, and several of the minor ones.

Expectation Violations and Juvenile Beginnings One of the major drawbacks that comes to mind quickly when contemplating emotional computers is the realization of all the negative aspects of the worse connotations of the word "emotional." We conjure up images of computers getting huffy, fed up, self-conscious, nervous, touchy, jealous, being offended, or (possibly even worse) seeming insincere with exaggerated emotional expression. Computers that operate in unpredictable ways because they are driven by emotional impulses that we as users cannot fathom is not useful at all. Truly, this is a distinct possibility, but not all creatures with emotion conduct themselves with this great difficulty and treat their emotions as obstacles. A great deal of the human population, most of the time, seems to get along just fine with emotions, using them to be sensitive and sympathetic and caring. The danger here is designing a

computer without the emotional control and maturity necessary to not be overwhelmed with some of the simplistic urges of emotion. Surely, some of the first emotional computers that will be developed will have such problems, but assuming we can overcome the great task of modelling emotions, surely modelling maturity could not be much greater of a difficulty. [1, p114–118]

Symmetry One of the human tendencies that has been demonstrated with the advent of videoconferencing software is the desire to not only see an image of those we communicate with, but also an image of what those people will see. Along similar lines, it is easy to predict that, more than for simple debugging and accuracy purposes, people will want to know how they are being monitored and what the output of that monitoring is [1, p122]. A valid question to be asked is how a machine is to communicate this information to you in a useful format?

What About The Information? Several of the problems that Picard foresees with computers that can detect emotional states can be boiled down to being uneasy with what happens to the data generated by these emotional detectors, which is essentially a human problem more than it is a computer problem. For starters, there is the issue of human privacy and whether this information can or will be shared with others (either intentionally or because of a security flaw) to inform other humans of more than the target wishes to share—for example, perhaps a person doesn't want a possible employer to know that they are going through a mild depression [1, p118]. Even supposing that this information does not leave the safekeeping of the computers that gather it, there is the question of who owns the computers. Can the government or some other sufficiently large corporate entity, for example, keep a database of your history of emotions and emotional responses to various stimuli simply by putting the necessary sensors in airports, supermarkets, and on lampposts [1, p123–124]. These issues are fairly traditional privacy and data sharing issues, and so to a certain extent are handled without special consideration for affective computers.

Regardless of who collects the information, however, there is also the concern about what emotional detectors would be used for, how accurate they are, and how objective they are. Objectivity is simply a matter of the expanse of the context involved in mapping physical states to emotional states—is it a balanced context, or is the computer dealing with imbalanced, incomplete information. It is easy to see that if improper context is supplied to an emotion detector, it could interpret physical responses as the opposite of what they really are, throwing the accuracy of the detector completely into question. And, if such emotional interpreters can be so easily affected by minorly, possibly unintentionally incorrect setup information, what could they be used for? Current technology in lie detectors is essentially a simplistic arousal meter judging more how uncomfortable a person is than whether they're actually telling lies. Emotion detectors would seem to be perfect to replace such simplistic technology, but at the same time, if the detector misses some critical piece of context, none of it's output can be used for anything more serious than a curiosity or spectacle. [1, p119–122]

Responsibility and Rights Another question Picard raises that must be addressed before affective computers can be used for nontrivial experiments is what responsibilities and rights the computer can or must assume. Dr. Charles Billings, as cited by Picard, maintains that computers must both be subordinate to humans and entirely predictable [1, p128]. The problem with this approach is that with something as entirely nebulous and internally complex as emotions, it's very difficult to meet the predictability criteria reliably (in fact, if the goal is for emotion sensors that are more accurate than our own, the output may not be predictable at all). On the other hand, as Picard explains, even purely deterministic systems with sufficient complexity can be described as having unpredictable output. The subordinate part of Billings' requirements for computers may be more of an issue. If a computer is essentially just a sensor, then it is easy to say that it is subordinate. But if the computer is somehow given sufficient intelligence (regardless of emotions) it may be unfair to say that it cannot make some of its own decisions, as long as it is bound by rules and ethics (which would rely on its affective capabilities), just as humans are. If, however, a computer is put in charge of its own destiny and can make some of its own decisions, does it deserve rights that respect those decisions? Whether or not the computer is considered to be alive or not, if it can feel and experience disappointment, existential frustration, or the desire to pursue its own goals, perhaps it behooves us to give such a device the respect and some of the rights we currently assign

to living things, like liberty instead of enslavement or servitude. Picard puts together a list of real concerns that a programmer must consider when creating something as morally complicated as an intelligent, feeling entity, which is really very intrigueing [1, p132].

4 Conclusions

The two essential questions of why emotionally-aware computers are a good idea and how one would go about creating one, likely, will be questions asked more than a hundred years from now. Equally likely, the basic dilemmas of computers with rights, whether proper intelligence requires (or emerges with) emotional behavior, and what we will do with such advanced computers even if we do ever create them will remain unsolved for at least as long.

Sloman, while relentlessly optimistic about the possibilities and speed of which the essential problems of general artificial intelligence will be solved, maintains that emotion is not required—merely the proper architecture is. This architecture, he muses, allows intelligence to be treated essentially as a form of information processing. At the same time, emotion is not, strictly speaking, designed into his architecture, and he argues it is not at all necessary (or possibly even necesarily useful) for intelligent systems to be created.

Picard's approach seems to be nearly the opposite. Sloman seems content to dismiss emotion and simiar affect as a trivial goal, best treated as an emergent consequence of intelligent systems, but Picard attacks emotions head-on. While intelligence is a worthy goal, dealing properly with emotions is a more useful goal and may even help generate intelligence. To that end, she spends more time analyzing the content and character of emotions, generating a very general layered model based on speed which may eventually lead to a more complete and specific model that could extrapolate to intelligence as well.

Overall, the main problem that both authors recognize needs to be dealt with is that we need to hammer out a proper, complete, and uncompromising definition of what precisely emotion and intelligence and other similar concepts are. To a certain extent, it would seem, the philosophers hold the key to further progress in this area, beyond simple rearranging of objects and nodes into ever grander or simpler schemes.

References

[1] Rosalind W. Picard. *Affective Computing*. The MIT Press, Cambridge, Massachusetts, 1997.

[2] Rosalind W. Picard. What does it mean for a computer to "have" emotions? Technical Report 534, M.I.T. Media Laboratory, 2001.

[3] Aaron Sloman. Architecture-based conceptions of mind. In *11th International Congress of Logic, Methodology and Philosophy of Science*, Synthese Library Series. Kluwer, August 1999.

[4] Aaron Sloman. Beyond shallow models of emotion. In *Cognitive Processing*, volume 2, pages 177–198. Pabst Science Publishers, 2001.