Emotionware

by **Lynellen D.S. Perry**

Intelligent agents seem to be the latest fad in artificial intelligence. Though no one seems to agree on exactly what is meant by the term ``agent,'' there are a growing number of them on the market. They are packaged under names like ``personal assistant,'' ``knowbot,'' ''softbot,'' ``userbot,'' ``knowledge navigator,'' and ``guide.'' The selling points for these programs are that they claim to have captured the essential qualities of human intelligence. According to Bates, this means ``reasoning, problem solving, learning via concept formation, and other qualities apparently central to the capacity we call intelligence'' **[1]**. Yet Bates claims that these qualities represent the ``idealized scientist,'' not the average human. To have a believable agent, one that seems to be alive, to think, and to feel, Bates asserts that there needs to be a display of emotion **[1]**. Emotionless agents are viewed as merely machines. This statement is highly controversial and this article will not delve into the mire of trying to fully explain each side of the argument. Instead, we will consider why emotions might be useful to have, and touch on what would be involved in developing the emotional side of an intelligent agent.

The agent technology is being researched to ``engage and help all types of end users'' **[2]** with such tasks as ``information filtering, information retrieval, mail management, meeting scheduling, selection of books, movies, music, and so

forth'' **[4]**. Agents that display emotion would certainly be beneficial in certain contexts. If humans identify with and accept an agent as human instead of machine-like, they may be more able to trust the agent and better able to communicate with it. This type of agent could then be a personal assistant, a companion to shut-ins, a counselor, or even a nurse who actually listens to your concerns and attempts to explain things to you and comfort you. But giving a program more than a rudimentary imitation of emotion will not be an easy thing. We don't even understand how human emotions work. This, however, brings us right to the point that solving this problem is a perfect AI task. By trying to model emotions, perhaps we can learn more about them. By learning more about them, we can create more realistic models for use by the agents.

The Challenge

In building emotional intelligent agents, Maes feels there are two major problems to be tackled.

> The first problem is that of *competence:* how does an agent acquire the knowledge it needs to decide when to help the user, what to help the user with and how to help the user? The second problem is that of *trust:* how can we guarantee the user feels comfortable delegating tasks to an agent? **[4]**

The problem of trust is where emotion comes into the discussion. Note that science fiction has been a primary definer of what intelligent agents ``should'' be. For example, the crux of the story line in ***Do Androids Dream of Electric Sheep?*** by Philip P.K. Dick is the controversy about the humanness of the replicants. Replicants are not metal machines, but biologically engineered beings 'born' as adults and assigned to various duties that are hazardous or undesirable for humans to perform. The newest generation of replicants have a new feature…fake memories have been implanted at their 'birth' so that they believe they have a past, a family, etc. These replicants are the first to be able to show emotion, and now it is hard for the main character in the story to kill them, which is his job. Before emotions, they were 'clearly' not quite human. They were merely synthetic soldiers and slaves that needed to be 'retired'. People have a ``tendency to anthropomorphize -- to see human attributes in any action appearing in the least bit intelligent'' **[5]** so we seem to be pre-disposed to thinking of a competent intelligent agent as being somewhat human even without the aid of an emotional response from the agent. But just as we feel a need to understand

and control a human secretary or administrative assistant, we also feel a need to be in control of the computational systems that are acting on our behalf.

Paralleling the logic of the above story, Bates says the ``appropriately timed and clearly expressed'' [1] display of emotion appears to be crucial to our ability to relate to a character or agent. If the end user can relate to the software agent, and understand its actions, they will feel more in control of the agent and thus more comfortable with letting it handle personal tasks. Bates suggests that there are three guidelines to follow for proper portrayal of emotion in a non-animate agent. First, ``the emotional state of the character must be clearly defined''[1]. This is not as simple as it may sound. Humans have a very wide range of emotions. For example, many self-help books have long lists of emotions to help the reader identify which they are feeling.

Emotions are not something easily defined and categorized, but that doesn't mean we shouldn't try. **Marvin Minsky** , a founding father of artificial intelligence, has a lot to say about categorizing emotions:

> Our culture has come to use the word 'emotion' for just about every kind of thinking or acting that we don't understand yet. Somehow, this labeling seems to have taken on a life of its own, so that it has become not scientifically respectable even to try to understand how they work. We've so often heard that emotions are mysterious that we've come to think that this is the way it should be [3].

So how does the software behind an intelligent agent decide which emotion to display? Bates' second guideline is that ``the thought process reveals the feeling''[1]. This is the approach taken by an electronic mail agent reviewed by Pattie Maes:

> The agent communicates its internal state to the user via facial expressions . . . The faces have a functional purpose: they make it possible for the user to get an update on what the agent is doing ``in the blink of an eye.'' There are faces for . . . ``unsure'' (the agent does not have enough confidence in its suggestion) . . . The ``pleased'' and ``confused'' face help the user gain information about the competence of the agent [4].

Expressing emotion to convey the thought processes of the agent ``helps us know that characters really care about what happens in the world, that they truly have desires. In contrast, the fighting characters of current video games show almost no reaction to the tremendous violence that engulfs them'' [1]. Riecken [3] suggests that one way to implement emotions in an intelligent agent would be to encode some ``primitive type of instinct-goals'' and then put the agent into the world to learn from experience. Over time, ``those instincts [might] become associated with experiences in that culture, and get dis-colored and reshaped by other concept-words used by humans, such as happiness or sadness or anger'' [3].

The third guideline proposed by Bates is to ``accentuate the emotion. Use time wisely to establish emotion, to convey it to viewers, and to let them savor the situation''[1]. The point here is that the temporal relation between events and the display of emotion is what helps other people understand what emotion is being displayed. Happiness during a fight, or yelling during a peaceful scene confuses people and causes them to not identify with the agent.

One possible way to encode emotions would be to present descriptions of situations along with the appropriate emotional response. This sort of a **case-based** approach would be coupled with other AI techniques such as reasoning by analogy. Once the emotional responses are learned, the program will still need a way to communicate those emotions. Should it use facial expressions, words, or some other method? Though 55% of human communication consists of body language, facial expressions can be ambiguous unless you are trained to read them carefully. Worse, body language can be culture dependent. Words are more precise, but will also require the user to have a larger vocabulary. For example, what exactly is the difference between deserted and abandoned? And do those words mean you feel angry or that you feel afraid, or both? So words are not necessarily any more clear than simulated facial expressions.

Another area of research is the question of how wide a range of emotions agents should possess. Boden [6] gives the following example:

> When Alice had finished reciting *You are old, Father William*, at the Caterpillar's request, the following exchange ensued:
> ``That is not said right,'' said the Caterpillar.
> ``Not quite right, I'm afraid,'' said Alice, timidly: ``Some of the words have got altered.''

>   ``It is wrong from beginning to end,'' said the Caterpillar decidedly.

No one would choose the Caterpillar as an assistant; he was both grumpy and unhelpful. What Alice wanted to know was just where she had gone wrong. . . but the Caterpillar was in no mood to tell her.

Just as when we are attempting to build any artificial intelligence, we have to ask ourselves how closely we want to model humans. Do we want to model the mistakes they make, or only their perfections? In this case, we have to ask if we want to endow the agent with the full range of emotions, which includes anger and grumpiness, or only the pleasant ones like happiness and compassion. And if the agent only portrays the pleasant emotions, will we still perceive it to be realistic and believable? Sometimes it seems that humans who are happy all the time are pretty unreal (and at times very annoying because of this quality). How much more would that apply to a computerized agent?

If we do create an agent with a full complement of emotional capability, along with human-like intelligence and ability to learn over time, will we discover that some agents develop emotional instabilities just like humans? Will we need to send some agents to a professional psychiatrist or will we just delete those agents and try again? On the other hand, an agent that has a bi-polar personality or psychosis might be a fantastic research tool to help us understand and treat humans with these disorders.

But what about the irrational behavior that occurs before diagnosis is made? Just as with a human, an agent who is trusted to perform sensitive, private, or important tasks while suffering from an emotional disorder may make a huge mess of things. We are quite comfortable blaming a human for the acts they commit, but who will be blamed when a psychotic agent commits a crime? The agent? The programmer who wrote the algorithm that the agent follows?

Conclusions

The capability of displaying emotion seems to be a critical component of creating intelligent agents with whom humans can comfortably relate and communicate. The emotional aspect distinguishes a dead machine from an agent who is believable, alive, and trustworthy. But it will be difficult to model human emotions accurately since there exists such a broad range with fine nuances of meaning. In addition, if we are able to model emotions, there are several practical and ethical questions that arise. Perhaps

these need to be considered carefully before we attempt to create this technology. Just because we can do a thing does not necessarily mean that we should do that thing.

References

1

Bates, Joseph. The Role of Emotion in Believable Agents. *Communications of the ACM* 37,7 (July 1994), 122-125.

2

Riecken, Doug. Intelligent Agents. *Communications of the ACM* 37,7 (July 1994), 18-21.

3

Riecken, Doug. A Conversation with Marvin Minsky About Agents. *Communications of the ACM* 37,7 (July 1994), 23-29.

4

Maes, Pattie. Agents that Reduce Work and Information Overload. *Communications of the ACM* 37,7 (July 1994), 31-40.

5

Norman, Donald A. How Might People Interact with Agents. *Communications of the ACM* 37,7 (July 1994), 68-71.

6

Boden, Margaret A. Agents and Creativity. *Communications of the ACM* 37,7 (July 1994), 117-121.