

# Artificial Emotions\*

Graeme Taylor<sup>†</sup>

December 5, 2004

Until fairly recently it was considered that emotions hindered, rather than benefitted, rational thought or activity. This was not always the case however, and arguments have been advanced for the consideration of emotion as a vital part of our decision making process. One such view is that were emotions redundant, we would never have evolved a capacity for them. Whilst its clear that a basic fear response offers an evolutionary advantage, to what end do we experience guilt?

Interest in this area has increased with the emergence of the field of affective computing- creating machines with artificial emotions. This is another example of a symbiotic relationship between research in AI/A-Life and biology/psychology: to implement convincing emotions in machines requires an understanding of our own, whilst considering the benefits of doing so may help to explain the emergence of emotion in humans in the first place. For instance, guilt can be seen as purely a hindrance for a lone human- but we evolved as social creatures, and here guilt is intimately tied up with trust. A capacity for guilt could motivate acting in the interest of the group rather than the individual, and its outward display confirm our trustworthiness to that group. In the context of the Prisoner's Dilemma, for instance, an agent which has to factor in guilty feelings might therefore have to re-evaluate the pay-offs, and then be more likely to co-operate with agents it considers likely to have the same problem. In this way, a negative emotion for the individual is advantageous to the group as a whole.

Thus if emotions are beneficial for humans, it follows that creating machines with an emotive capacity may be helpful for our interaction with them, and more questionably for the machines' interaction with the environment. There are also some interesting issues that arise at a philosophical/religious level. I'll look at all three areas in turn.

---

\*First appeared on Everything2, at [http://www.everything2.com/index.pl?node\\_id=1687559](http://www.everything2.com/index.pl?node_id=1687559)

<sup>†</sup>Once again, this is a writeup inspired by a millenium lecture, entitled Can Robots Have Emotions, by Dylan Evans (title shamelessly stolen for part of this work). His personal website is at <http://www.dylan.org.uk/index.html> where you can find links to the IAS. He's written a book, *Emotion- the science of sentiment* (ISBN 0-19-285376-7) which I read; a modified version of the chapter on artificial emotions can also be found on his site.

## Interacting with Intelligent Machines

If AI research has a difficult task in convincing its critics of its validity then artificial emotions have an even more daunting task. This is because the very fact that we desire genuine emotion in others leads us to naturally question their sincerity lest we be exploited by an unscrupulous other. The importance of distinguishing a real emotion from a faked one has far more relevance to our day to day lives than debating whether the person opposite us is really conscious or just behaving in a consistent manner with our own experience of consciousness.

For robotics, this psychological baggage crops up as a phenomenon known as the Uncanny Valley- the assertion that a 90% life-like robot is more disconcerting than one that is barely human-like at all. This can be explained by a willingness to anthropomorphise clearly-non human animals or machines, ascribing human behaviour and feeling to them whilst not attributing enough malice or complexity for that behaviour to be less than genuine; whilst a close approximation to a human might trigger our doubts about sincerity, all the more so for its mechanical underpinnings. This is not just a problem in robotics- in the realm of 3d rendering, it seems that the bright green creations of Shrek 2 are somehow more believable than the painstaking-but-not-quite-right characters of the Final Fantasy movie. Nor is it a new idea- a fear of the inanimate becoming animate (Frankenstein's monster, zombies, demonic possession) is a common theme in horror and gothic fiction.

Ultimately, however, artificial emotions will probably play a vital role in bridging the Uncanny Valley to truly convincing robotics. Emotions are the most commonly advanced factor for humans being distinct from machines, outranking more heavily researched areas such as autonomy or intelligence.

But beyond acceptance by human users, or understanding our own feelings, what other motives are there for creating machines with an emotional capacity? A computer with an ability to detect stress in its user would have applications from the mundane (offering more pleasant emails first to a frazzled office worker) to the life-saving (assisting fighter-pilots). But the biggest area may be recreational. From the Sims to Sony's AIBO robots, there's an increasing market for toys that seem to feel. The market may therefore provide the litmus test for convincing emotions- even considering the valley of the uncanny it's hard to find much other explanation for the success of those hideous Furbys than their status as a must-have Christmas gift.

## Can Robots have emotions?

Having established the desirability of emotions in machines, is there any hope of creating them? This, of course, becomes a question of semantics. One view is that humans are no more than robots themselves (more on this later), and we have emotions- so clearly, emotional robots are possible. The challenge then becomes to create artificial (as opposed to natural) robots which

share this ability. At the other extreme is the belief that emotion is only meaningful as a conscious experience, not a simulated behaviour, and this, coupled with a belief that machines cannot be conscious entities, would rule out a machine from ever having 'true' emotion.

The view of emotion within the field of psychology/philosophy is slightly different from the casual description of a feeling- it is held that the experience of emotion is tied to a number of aspects, encompassing reflex action, higher thought processes, facial expressions, biochemical responses and so on- which need not all be present. We wouldn't deny someone the capacity for joy simply because facial paralysis prevents them smiling- so why say a robot needs a full set of human biological structures to express happiness?

In fact, one of the most well known projects in the area of affective computing is Kismet[1]- a robot that generates human facial expressions to convey emotions, developed at MIT. When video samples (available online) of Kismet were shown at a lecture I attended, the audience was quick to empathise with the intended emotion- especially when it seemed sad. In the UK, interest in robotics (fuelled by shows such as Robot Wars or the exhibitions at MAGNA) has also led to the establishment of departments such as at the University of the West of England's Intelligent Autonomous Systems Laboratory (IAS). Commercially, AIBO is described as having six different emotional states, whilst the addictive nature of the Tamagotchi or other virtual pets is testament to our willingness to ascribe human emotions onto computer simulations, or feel a need to nurture them.

## Philosophical Implications

If we do succeed in creating genuinely emotional robots, then this (alongside artificial intelligence) is seen by many as a challenge to our status as anything more than machines ourselves, since it removes a means of distinguishing artificial from natural.

Indeed, the early (13th-18th century) history of attempts at lifelike machines is littered with examples of would-be roboticists being persecuted for witchcraft, sacrilege or heresy; usually their machines get destroyed in the process. The challenge to religion of equating humanity, other animals and mechanical robots as being no more than machines is clear, but I feel that we'll always point to something that we see as making us 'special' or better.

But regardless of the challenge to our own humanity, robots with feelings raise their own special set of ethical questions, that have been explored more by science fiction than science proper. The most famous must be the three laws of robotics, but there are many other examples, usually revolving around when things go wrong- from Bladerunner to The Terminator. In Descartes' time, it was argued that animals were not capable of 'really' experiencing pain- that their apparent anguish was just as mechanical a response as our own reflexes. These days most would accept

that its possible to commit cruelty to animals (there are even laws against it), so how long after emotional machines would a robot rights movement emerge?

## References

- [1] Information on Kismet, including pictures and video, is at  
*<http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html>*