Opinion: AI like HAL 9000 can never exist because real emotions aren't programmable

April 9 2018, by Guillaume Thierry

HAL 9000 is one of the best-known artificial intelligence characters of modern film. This [superior form of sentient computer](#) embarks on a mission to Jupiter, along with a human crew, in Stanley Kubrick's iconic film 2001: A Space Odyssey, which is currently celebrating its 50th year since release.

HAL is capable of speech production and comprehension, facial recognition, lip reading – and [playing chess](#). Its superior computational

ability is boosted by uniquely human traits, too. It can interpret emotional behaviour, reason and appreciate art.

By giving HAL emotions, writer Arthur C. Clarke and filmmaker Stanley Kubrick made it one of the most human-like fictional technologies ever created. In one of the most beautiful scenes in sci-fi history, it says it is "afraid" when mission commander Dr. David Bowman starts disconnecting its memory modules following a series of murderous events.

HAL is programmed to deliver optimal assistance to the crew of the spaceship Discovery. It has control over the entire vessel, and staggering intelligence to aid it in its task. Yet soon after we become acquainted with HAL, we cannot help feeling that it is worried – it even claims it is experiencing fear – and that it has an ability to empathise, however small. But while there is nothing to preclude the *idea* that such an emotional AI could see the light of day, if such depth of feelings were to be included in real world technology, they would have to be entirely fake.

A 'perfect' AI

When, during the film, Bowman starts to manually override HAL's functions, it asks him to stop, and after we witness a fascinating obliteration of HAL's "mental" faculties, the AI seemingly tries to comfort itself by singing Daisy Bell – reportedly the first ever song produced by a computer.

In fact, viewers begin to feel that Bowman is killing HAL. The disconnection feels like a vengeful termination, after witnessing the film's earlier events. But though HAL makes emotional statements, a real world AI would certainly be limited to having only the ability to reason, and make decisions. The cold, hard truth is that – despite what

computer scientists say – we will never be able to program emotions in the way HAL's fictional creators did because we do not understand them. Psychologists and neuroscientists are certainly trying to learn how emotions interact with cognition, but still they remain a mystery.

Take our own research, for example. In a study conducted with Chinese-English bilinguals, we explored how the emotional value of words can change unconscious mental operation. When we presented our participants with positive and neutral words, such as "holiday" or "tree", they unconsciously retrieved these word forms in Chinese. But when the words had a negative meaning, such as "murder" or "rape", their brain blocked access to their mother tongue – without their knowledge.

Reason and emotion

On the other hand, we know a lot about reasoning. We can describe how we come to rational decisions, write rules and turn these rules into process and code. Yet emotions are a mysterious evolutionary legacy. Their source is the source of everything, and not simply an attribute of the mind that can be implemented by design. To program something, you not only need to know how it works, you need to know what the objective is. Reason has objectives, emotions don't.

In an experiment conducted in 2015, we were able to put this to the test. We asked native speakers of Mandarin Chinese studying at Bangor University to play a game of chance for money. In each round, they had to take or leave a proposed bet shown on the screen – for example, a 50% chance of winning 20 points, and a 50% chance of losing 100 points.

We hypothesised that giving them feedback in their mother tongue would be more emotional to them and so lead them to behave differently, compared to when they received feedback in their second

language, English. Indeed, when they received positive feedback in native Chinese, they were [10% more likely to take a bet in the next round](), irrespective of risk. This shows that emotions influence reasoning.

Going back to AI, as emotions cannot be truly implemented in a program – no matter how sophisticated it may be – the reasoning of the computer can never be changed by its feelings.

One possible interpretation of HAL's strange "emotional" behaviour is that it was programmed to simulate emotions in extreme situations, where it would need to manipulate humans not on the basis of reasoning but by calling upon their [emotional]() self, when human reason fails. This is the only way I can see that real world AI could convincingly simulate emotions in such circumstances.

In my opinion, we will not, ever, build a machine that feels, hopes, is scared, or happy. And because that is an absolute prerequisite to any claim that we have engendered artificial general intelligence, we will never create an artificial mind outside life.

This is precisely where the magic of 2001: A Space Odyssey lies. For a moment, we are led to believe the impossible, that pure science fiction can override the facts of the world we live in.

This article was originally published on [The Conversation](). Read the [original article]().

Provided by The Conversation