Too close for comfort

TIMOTHY ARCHIBALD

We often get creeped out by artificial beings that look like us, says Joe Kloc, but the reasons are more complex than you might imagine

EIGHT years ago, Karl MacDorman was working late at Osaka University in Japan when, around 1 am, his fax machine sputtered into life. Out came a 35-year-old essay, written in Japanese, sent by a colleague.

It was an intriguing read for MacDorman, who was building hyperrealistic androids at the time. It warned that when artificial beings have a close human likeness, people will be repulsed. He and his colleagues worked up a quick English translation, dubbing the phenomenon the "uncanny valley".

They assumed their rough draft of this obscure essay would only circulate among roboticists, but it caught the popular imagination. Journalists used the uncanny valley to explain the lacklustre box office performance of movies like *Polar Express*, in which audiences were creeped out by the computer-generated stars. It was also blamed for the failure of humanoid robots to catch on. Finding an explanation for why the uncanny valley occurs, it seemed, would be worth millions of dollars to Hollywood and the robotics industry. Yet when researchers began to study the phenomenon, citing MacDorman's translation as the definitive text, answers eluded them.

MacDorman now believes we have been looking at the uncanny valley too simplistically, and he partly blames his own rushed translation. He and others are converging on an explanation for what's actually going on in the brain when you get that uncanny feeling. If correct, the phenomenon is more complex than anyone realised, encompassing not only our relationship with new technologies but also with each other.

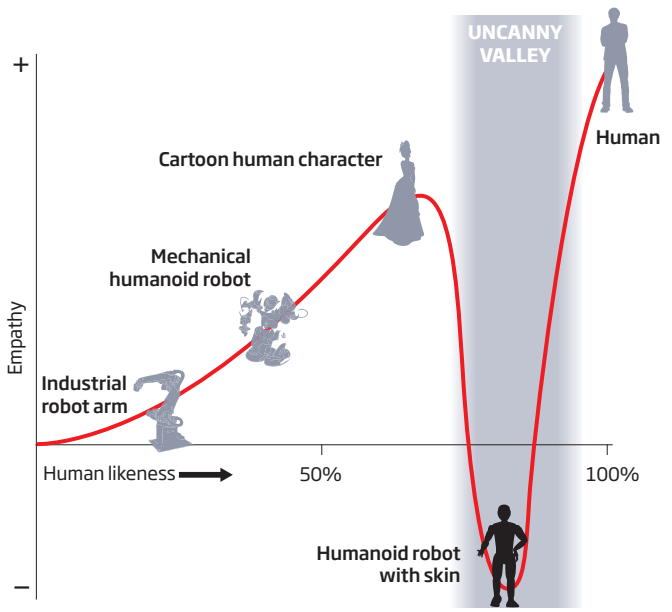
While it's well known that abnormal facial and body features can make people shun others, some researchers believe that human-like creations unnerve us in a specific way. The essay that MacDorman read was published in 1970 by roboticist Masahiro Mori. Entitled "*Bukimi No Tani*" – or The Valley of Eeriness – it proposed that humanoid robots can provoke a uniquely uncomfortable emotion that their mechanical cousins do not.

For decades, few outside Japan were aware of Mori's theory. After MacDorman's translation brought it to wider attention, his ideas were extended to computer-generated human figures, and research began in earnest into the uncanny valley's possible causes.

MacDorman's first paper on the subject examined an idea proposed by Mori: that we feel uncomfortable because almost-human ›

Losing empathy

When artificial creatures look like us, they enter the "uncanny valley". This could be because we feel less empathy towards them, as this illustration shows



CREDIT: WDISNEY/EVERETT/REX FEATURES

TIMOTHY ARCHIBALD

robots appear dead, and thus remind us of our own mortality. To test this, MacDorman used something called terror management theory. This suggests that reminders of death govern much of our behaviour – including making us cling more strongly to aspects of our own world view, such as religious belief.

So MacDorman asked volunteers to fill in a questionnaire about their world views after showing them photos of human-like robots. Sure enough, those who had seen the robots were more defensive of their view of the world than those who had not, hinting that the robots were reminding people of death.

This explanation makes intuitive sense, given that some animated characters and robots appear corpse-like. But even at the time it was clear to MacDorman that the theory had its limits: reminding someone of their own demise does not, on its own, elicit the uncanny response people describe. A gravestone reminds us of death, for example, but it doesn't make us feel the same specific emotion.

Competing theories soon emerged. Some researchers blamed our evolutionary roots; we have always been primed to shun unattractive mates, after all. Others blamed the established idea that we evolved feelings of disgust to protect us from pathogens. Christian Keysers of the University of Groningen in the Netherlands pointed out that irregularities in an almost-human form make it look sick. Since uncanny robots look very similar to us, he argued, we may subconsciously believe we are at a higher risk of catching a disease from them.

Again, both these theories are incomplete: many disgusting and unattractive things do not, by themselves, elicit that specific uncanny feeling. We know that somebody sneezing on the subway exposes us to potentially dangerous pathogens, yet a subway ride is not an uncanny experience. "There are too many theories," says MacDorman. "The field is getting messy, further away from science."

The first clue there was something more complex going on came when neuroscientists began to explore what might be happening in the brain. In 2007, Thierry Chaminade of the Advanced Telecommunications Research Institute in Kyoto, Japan, and colleagues presented people with a series of computer-generated characters that resembled humans to varying degrees, while monitoring their brain activity in an fMRI machine. While it wasn't the specific aim of the study, the results hinted at a new explanation for the uncanny. When the volunteers observed a character that appeared almost human, activity increased in the part of their brain responsible for mentalising – the ability to comprehend the mental state of another.

> "The uncanny feeling could be caused by a dissonance in the system that helps us feel empathy - the ability to feel what something else feels"

Mentalising is understood to be involved in feeling empathy. Could empathic pathways in the brain be responsible for mediating the uncanny response?

More evidence came in 2011 with a second fMRI study, this time led by Ayse Saygin at the University of California, San Diego. The researchers observed people's brain activity while showing them video footage of a mechanical robot, a human and a lifelike android known to induce the uncanny valley response. Each of these were displayed to the participants performing an identical action – but one triggered a notably different result.

When people observed the human or mechanical robot walking, the brain exhibited very little activity. But when participants had to process the lifelike android doing the same action, activity increased considerably in the visual and motor cortices of their brains.

Saygin and colleagues suggested that the feelings of eeriness produced by watching the android may stem from the extra work the brain needs to do to reconcile the robot's movements with the human-like behaviour it expects based on appearances.

It is thought that the motor cortex houses mirror neurons, which are specialised for a particular task and can also fire when we see another organism performing that task. While opinion remains divided on their role, these neurons have also been implicated in our ability to empathise with others.

The uncanny feeling, then, could be caused by a sort of dissonance in the system that helps us to feel empathy, says MacDorman (see illustration, above). "It seems related to the ability to feel what something else feels."

Mori didn't know this when he wrote his essay in 1970, but he did leave the door open to the possibility. When MacDorman translated the essay into English, he made a crucial simplification. According to the 2005 translation, when we are in the uncanny valley, our feelings of "familiarity" plummet. This quality – along with "likeability" – has provided the framework for countless studies of the uncanny valley, says MacDorman – and this may have been obscuring its possible roots in empathy.

Suppressed empathy

Mori didn't actually use the terms familiarity or likeability. Instead he used a neologism, *shinwakan*, which he invented because there was no opposite to the word uncanny. MacDorman now believes that *shinwakan* is actually a form of empathy. Last June, he published a new translation that he hopes will prompt researchers to look at the uncanny valley through this lens instead.

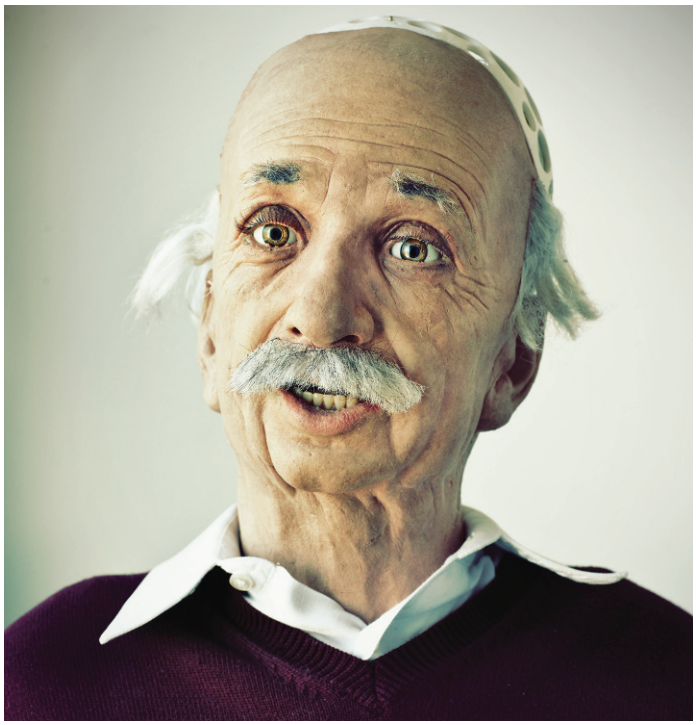
"The fact that empathy is complex means we can tease it apart," he says, "and figure out what is really at play."

In cognitive neuroscience, empathy is often divided into three categories: cognitive, motor and emotional. Cognitive empathy is essentially the ability to understand another's perspective and why they make certain decisions – to play "social chess", as MacDorman puts it. Motor empathy is the ability to mimic movements like facial expressions and postures, and emotional empathy is essentially sympathy, or the ability to feel what others feel. MacDorman's theory is that the uncanny feeling is produced when we experience certain types of empathy but not others. "The question," he says, "is what kind of empathy is being suppressed?"

To test one possibility, MacDorman, now at Indiana University in Indianapolis, asked people to watch videos of robots, computer-generated characters and real people in situations ranging from harmless to harmful. He then asked the volunteers to categorise

TIMOTHY ARCHIBALD

With close likenesses, it only takes a hint of the non-human to create an eerie feeling



TIMOTHY ARCHIBALD

these characters as either happy or sad about their situations. In other words, he was measuring participants' abilities to sympathise with the figures.

MacDorman found that they had a more difficult time determining the emotional state of characters that fell within the uncanny valley. This was, he believes, an indication that emotional empathy was being suppressed. On a cognitive and motor level, all the typical cues for empathy are triggered, but we can't muster sympathy, he says.

Kurt Gray, a psychologist at the University of North Carolina, Chapel Hill, agrees that the uncanny valley is about our inability to feel certain types of empathy, and that we should start looking at the phenomenon differently. "What Karl did in terms of framing is really important," he says.

Gray believes he has an explanation for why struggling to sympathise with human-like robots and animated characters would make us uncomfortable. In a recent study, he and Daniel Wegner at Harvard University asked volunteers to take a survey that measured their comfort level with various types of computer capabilities. The idea was to identify which human traits, when exhibited by a machine, make people uncomfortable.

The pair found that people thought computers capable of feeling emotions were the most unnerving. "We are happy to have robots do things, but not feel things," they concluded.

Gray's argument is that almost lifelike robots make us feel uneasy because we see in them the shadow of a human mind, but one that we know we can never comprehend. In other words, it's not just about our failure to sympathise with uncanny robots and computer-generated characters; it's also about our perception that they can empathise with us.

The particular brand of sympathy we reserve for other people requires us to believe the thing we are sympathising with has a self. And this concession of a mind to something not human makes us uncomfortable.

It follows that as long as we are aware that a robot or virtual character is not human, we will never grant it passage to cross the uncanny valley. Even if we do find a way to make artificial creatures with identical human features, they may still provoke discomfort if we know they are not like us. This possibility has already been explored in science fiction: consider how the human characters reacted to the cylons in *Battlestar Galactica*, says roboticist Christoph Bartneck of the University of Canterbury in New Zealand. "You have these robots indistinguishable from humans. That was what's so scary. They are not like us. But they are like us."

Perhaps this is what Mori was getting at when, years after he penned his essay, a reporter asked him if he thought humankind would ever build robots that crossed the uncanny valley: "Why try?" he responded.

The idea that the uncanny valley may be impossible to cross may come as bad news to Hollywood and robot designers. But it also stands as a sign of something many will find reassuring: that there is a particular feeling of empathy that only humans can share. ■

Joe Kloc is a writer based in New York