'Uncanny Valley': Brain network evaluates robot likeability

July 1 2019

Images of robots shown to participants, from least human-like (left) to most human-like (right). Credit: Rosenthal-von der Pütten *et al.*, *JNeurosci* 2019

Scientists have identified mechanisms in the human brain that could help explain the phenomenon of the 'Uncanny Valley' - the unsettling feeling we get from robots and virtual agents that are too human-like. They have also shown that some people respond more adversely to human-like agents than others.

As technology improves, so too does our ability to create life-like artificial agents, such as robots and computer graphics—but this can be a double-edged sword.

"Resembling the human shape or behaviour can be both an advantage and a drawback," explains Professor Astrid Rosenthal-von der Pütten, Chair for Individual and Technology at RWTH Aachen University. "The likeability of an artificial agent increases the more human-like it becomes, but only up to a point: sometimes people seem not to like it when the robot or computer graphic becomes too human-like."

This phenomenon was first described in 1978 by robotics professor Masahiro Mori, who coined an expression in Japanese that went on to be translated as the 'Uncanny Valley'.

Now, in a series of experiments reported in the *Journal of Neuroscience*, neuroscientists and psychologists in the UK and Germany have identified mechanisms within the brain that they say help explain how this phenomenon occurs—and may even suggest ways to help developers improve how people respond.

"For a neuroscientist, the 'Uncanny Valley' is an interesting phenomenon," explains Dr. Fabian Grabenhorst, a Sir Henry Dale Fellow and Lecturer in the Department of Physiology, Development and Neuroscience at the University of Cambridge. "It implies a neural mechanism that first judges how close a given sensory input, such as the image of a robot, lies to the boundary of what we perceive as a human or non-human agent. This information would then be used by a separate valuation system to determine the agent's likeability."

To investigate these mechanisms, the researchers studied brain patterns in 21 healthy individuals during two different tests using functional magnetic resonance imaging (fMRI), which measures changes in blood

flow within the brain as a proxy for how active different regions are.

In the first test, participants were shown a number of images that included humans, artificial humans, android robots, humanoid robots and mechanoid robots, and were asked to rate them in terms of likeability and human-likeness.

Then, in a second test, the participants were asked to decide which of these agents they would trust to select a personal gift for them, a gift that a human would like. Here, the researchers found that participants generally preferred gifts from humans or from the more human-like artificial agents—except those that were closest to the human/non-human boundary, in-keeping with the Uncanny Valley phenomenon.

By measuring brain activity during these tasks, the researchers were able to identify which brain regions were involved in creating the sense of the Uncanny Valley. They traced this back to brain circuits that are important in processing and evaluating social cues, such as facial expressions.

Some of the brain areas close to the visual cortex, which deciphers visual images, tracked how human-like the images were, by changing their activity the more human-like an artificial agent became—in a sense, creating a spectrum of 'human-likeness'.

Along the midline of the frontal lobe, where the left and right brain hemispheres meet, there is a wall of neural tissue known as the medial prefrontal cortex. In previous studies, the researchers have shown that this brain region contains a generic valuation system that judges all kinds of stimuli; for example, they showed previously that this brain area signals the reward value of pleasant high-fat milkshakes and also of social stimuli such as pleasant touch.

In the present study, two distinct parts of the [medial prefrontal cortex](#) were important for the Uncanny Valley. One part converted the human-likeness signal into a 'human detection' signal, with activity in this region over-emphasising the boundary between human and non-human stimuli—reacting most strongly to human agents and much less to artificial agents.

The second part, the ventromedial prefrontal cortex (VMPFC), integrated this signal with a likeability evaluation to produce a distinct activity pattern that closely matched the Uncanny Valley response.

"We were surprised to see that the ventromedial prefrontal cortex responded to artificial agents precisely in the manner predicted by the Uncanny Valley hypothesis, with stronger responses to more human-like agents but then showing a dip in activity close to the human/non-human boundary—the characteristic 'valley'," says Dr. Grabenhorst.

The same brain areas were active when participants made decisions about whether to accept a gift from a robot by signalling the evaluations that guided participants' choices. One further region—the amygdala, which is responsible for emotional responses—was particularly active when participants rejected gifts from the human-like, but not human, artificial agents. The amygdala's 'rejection signal' was strongest in participants who were more likely to refuse gifts from artificial agents.

The results could have implications for the design of more likable artificial agents. Dr. Grabenhorst explains: "We know that valuation signals in these brain regions can be changed through social experience. So, if you experience that an artificial agent makes the right choices for you—such as choosing the best gift—then your ventromedial prefrontal cortex might respond more favourably to this new social partner."

"This is the first study to show individual differences in the strength of

the Uncanny Valley effect, meaning that some individuals react overly and others less sensitively to human-like artificial agents," says Professor Rosenthal-von der Pütten. "This means there is no one robot design that fits—or scares—all users. In my view, smart robot behaviour is of great importance, because users will abandon robots that do not prove to be smart and useful."

More information: Rosenthal-von der Pütten, AM et al. Neural Mechanisms for Accepting and Rejecting Artificial Social Partners in the Uncanny Valley. *Journal of Neuroscience*; 1 July 2019; DOI: 10.1523/JNEUROSCI.2956-18.2019

Provided by University of Cambridge