

# After confronting one uncanny valley, another awaits

Jan-Philipp Stein & Karl F. MacDorman



The ‘uncanny valley’ has guided robot engineers on the limits of human likeness, yielding design principles to mitigate the risk of creepy robots. Yet unease with advancements in AI has exposed a new ‘uncanny valley of mind’, with researchers now exploring acceptable boundaries on simulating human intelligence, emotion, empathy and creativity.

In 1970, when engineers were just unlocking the potential of creating robotic machinery in the likeness of humans, a Japanese robotics professor, Masahiro Mori, issued a warning: In his paper *Bukimi no tani genshō*<sup>1</sup>, he observed that equipping robots with distinctly humanlike features could backfire – especially if the machines ended up looking like a real person, but not entirely. Mori’s metaphor of an ‘uncanny valley’ that machines risked “falling into” before reaching perfect human likeness was never intended as a scientific theory but instead as a designer’s guideline. Nevertheless, it emerged as a popular model for academic research<sup>2</sup>, with its English translation cited thousands of times<sup>3</sup>, and its predictions scrutinized from various angles<sup>4</sup>.

One might say that understanding the creepiness of near-perfect human replicas has prompted a paradigm shift in modern-day robotics. Instead of striving for humanlike skin textures, photorealistic eyes, or fully anthropomorphic movements, robot designers now often settle for more stylized faces and bodies or subscribe to the universal appeal of the *Kindchenschema* – designs that adapt and exaggerate childlike features<sup>5</sup>. Moreover, Mori’s reflections on robotic uncanniness have been woven into the texture of popular culture. The unfavourable reception of computer-animated movies with creepily lifelike protagonists has taught film producers a lesson about the uncanny valley’s effect on box office sales. We owe it to Mori’s work and the many contributions that have built on it that robots and animated characters in mainstream media are rarely designed to look as human as possible, confined instead to being googly-eyed, charming cartoons.

None of this is to say that efforts have ceased to design human replicas that climb out of the uncanny valley on the human side. Designers and roboticists still aim to create humanlike simulations flawless enough to be embraced by the public. For the time being, however, the path taken is often cautious, heeding the warning that Mori expressed more than 50 years ago.

This is only the story’s first chapter. Why we like or dislike a robot is mainly rooted in its visual appearance. But people’s perceptions also depend on the mental abilities the robot seems to present – in other words, the ‘mind’ that we ascribe to it. Acknowledging this relationship, and the much-publicized triumphs of recent artificial intelligence (AI) technology, such as the large language model ChatGPT, scientific

research has uncovered a new barrier to harmonious human–robot interaction: an uncanny valley of mind<sup>6,7</sup>, encompassing people’s aversion to machines that behave, think, or even feel too humanlike, and thus are experienced as eerie.

Although present-day AI is still far removed from replicating the complex processes of the human brain – let alone the biological foundations of genuine human emotion – this may not matter to lay observers. The sheer impression of an autonomous computer system having emotions or empathizing with its human user may already feel unpleasant or inappropriate.

In reckoning with this response, recent studies have underscored threatened human distinctiveness as an explanation<sup>8</sup>. This threat may be fostered by common norms and philosophical worldviews, especially in Western societies. Most people share a sense of what makes our species unique. Abilities such as higher-order cognition, emotion, empathy and creativity have been interpreted as placing humans at the ‘crown of creation’, an interpretation with a three-millennia history in Abrahamic religions<sup>9</sup>. The prospect of suddenly losing this unique position in the hierarchy of beings (and things) may feel threatening to some – not only as a symbolic threat such as a loss of human identity or dominance, but also in immediate and practical ways, as new AI-based machines render an increasing number of human jobs obsolete<sup>10</sup>.

So, having arrived at the edge of a new uncanny valley, how should we proceed? The first step is to engage in honest and difficult discussions. For example, how will the new abilities of smart technology weigh against the uncanny valley of mind? Will generative AI such as ChatGPT serve to increase human creativity – or will it end as another example of human hubris? Can advanced military robots reduce casualties in global warfare, or will they cause yet more suffering and chaos?

There is a utopian vision of robots – as friendly companions for isolated older adults, as instruments for scientific research, even as replacements for corrupt or incompetent leaders<sup>11</sup>. But is this vision enough, given the consequences, the disorientation, perhaps to the point of losing our sense of what it means to be human? How deceptive and inauthentic will our relationships with technology become? How narcissistic will we become, as the people in our lives are replaced by ever more compliant machines<sup>12</sup>? Are we facing a societal transformation with an unpredictable, perhaps dire, outcome?

Certainly, unravelling these and similar questions will be daunting. It will require insight from numerous fields and some intellectual bravery. To avert catastrophe, we must work out machine ethics – a clear, well-defined understanding of what a machine should and should not do<sup>13</sup>.

Then again, considering humankind’s yearning for progress and pushing the limits of our species’ creativity, we might already be past that point. Even moral principles may be thwarted by financial and political interests. But there is time to develop a shared understanding of what we want our robots to be: our tools, our toys, our workers, or our companions – for one day, AI may become our equal or even our master.

Jan-Philipp Stein<sup>1</sup> & Karl F. MacDorman<sup>2</sup>  

<sup>1</sup>Department of Media Psychology, Chemnitz University of Technology, Chemnitz, Germany. <sup>2</sup>Luddy School of Informatics, Computing, and Engineering, Indiana University, Indianapolis, IN, USA.

 e-mail: [kmacdorm@iu.edu](mailto:kmacdorm@iu.edu)

## References

1. Mori, M. Bukimi no tani [The uncanny valley]. *Energy* **7**, 33–35 (1970).
2. Diel, A., Weigelt, S. & MacDorman, K. F. A meta-analysis of the uncanny valley's independent and dependent variables. *ACM Trans. Hum.-Robot Interact.* **11**, 1–33 (2022).
3. Mori, M., MacDorman, K. F. & Kageki, N. The uncanny valley [from the field]. *IEEE Robot. Autom. Mag.* **19**, 98–100 (2012).
4. Wang, S., Lilienfeld, S. O. & Rochat, P. The uncanny valley: existence and explanations. *Rev. Gen. Psychol.* **19**, 393–407 (2015).
5. Cho, S., Dydynski, J. M. & Kang, C. Universality and specificity of the *Kindchenschema*: A cross-cultural study on cute rectangles. *Psychol. Aesthet.Creat. Arts* **16**, 719–732 (2022).
6. Gray, K. & Wegner, D. M. Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition* **125**, 125–130 (2012).
7. Stein, J.-P. & Ohler, P. Venturing into the uncanny valley of mind – The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition* **160**, 43–50 (2017).
8. Cha, Y. et al. Compensating for the loss of human distinctiveness: the use of social creativity under human–machine comparisons. *Comput. Hum. Behav.* **103**, 80–90 (2020).
9. MacDorman, K. F., Vasudevan, S. K. & Ho, C.-C. Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *AI Soc.* **23**, 485–510 (2009).
10. Złotowski, J., Yogeewaran, K. & Bartneck, C. Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *Int. J. Hum. Comput. Stud.* **100**, 48–54 (2017).
11. Spatola, N. & MacDorman, K. F. Why real citizens would turn to artificial leaders. *ACM Digit. Gov. Res. Pract.* **2**, 1–24 (2021).
12. Turkle, S. Authenticity in the age of digital companions. *Interact. Stud.* **8**, 501–517 (2007).
13. Wallach, W. & Allen, C. *Moral Machines: Teaching Robots Right from Wrong* (Oxford Univ. Press, 2010).

## Competing interests

The authors declare no competing interests.