Subjective Ratings of Robot Video Clips for Human Likeness, Familiarity, and Eeriness: An Exploration of the Uncanny Valley

Karl F. MacDorman
School of Informatics, Indiana University, USA

Abstract

Masahiro Mori observed that as robots come to look more humanlike, they seem more familiar, until a point is reached at which subtle deviations from human norms cause them to look creepy. He referred to this dip in familiarity and corresponding surge in strangeness as the *uncanny valley.* The eerie sensation associated with a mismatch between human expectations and a robot's behavior provides a useful source of feedback to improve the cognitive models implemented in the robot. Is the uncanny valley a necessary property of near-humanlike forms? This paper contributes to ongoing work in understanding the nature and causes of the uncanny valley by means of an experiment: 56 participants were asked to rate 13 robots and 1 human, shown in video clips, on a very mechanical (1) to very humanlike (9) scale, a very strange (1) to very familiar (9) scale, and a not eerie (0) to extremely eerie (10) scale. Contrary to earlier studies with morphs [MacDorman and Ishiguro, 2006], plots of average and median values for ratings on these scales do not reveal a single U-shaped valley as predicted by Mori's uncanny valley hypothesis [1970], although his hypothesis allows for some variation owing to movement. Robots rated similarly on the *mechanical versus humanlike* scale can be rated quite differently on the *strange versus familiar* or the *eeriness* scales. The results indicate that the perceived human likeness of a robot is not the only factor determining the perceived familiarity, strangeness, or eeriness of the robot. This suggests that other factors could be manipulated to vary the familiarity, strangeness, or eeriness of a robot independently of its human likeness.

Introduction

To build robots that at least superficially approach human likeness is leading to insights in human perception and face-to-face interaction. These *android* robots possess the physical presence that simulated characters lack, yet can be more perfectly controlled than any human actor, to isolate the factor under study. Even in experiments in which the android's responses are identical, we can observe how human responses vary according to their beliefs. For example, Japanese participants showed the same modesty with their eyes by averting gaze downward when interacting with an android as when interacting with a human interlocutor *if* they believed the android were under human control by telepresence [MacDorman et al., 2005].

In addition, androids provide an ideal testing ground for theories from the social and cognitive sciences because competing models can be implemented in an android and then tested by letting the android interact with people [MacDorman and Ishiguro, 2006]. Androids provide not only a platform for integrating techniques from science and engineering but also for studying the relationship between interaction and cognitive mechanism. Thus, they may one day provide an avenue for unifying the behavioral sciences and cognitive neuroscience.

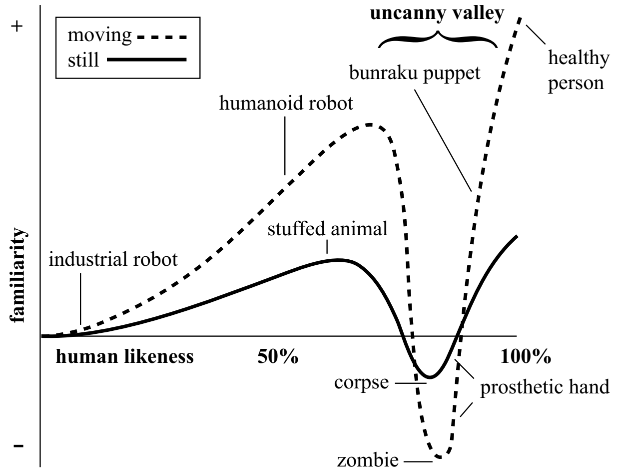

Figure 1: Mori hypothesized the relation between human likeness and perceived familiarity: familiarity increases with human likeness until an *uncanny valley* is reached caused by sensitivity to perceived imperfections in near-humanlike forms [Mori, 1970]. Movement, according to Mori, magnifies the uncanny valley.

Recent evidence indicates that androids are generally better able to elicit human-directed norms of interaction than less humanlike robots or animated characters [MacDorman and Ishiguro, 2006] [Cowley and MacDorman, 2006]. However, Mori [1970] observed a heightened sensitivity to defects in near-humanlike forms—an *uncanny valley* in what is otherwise a positive relationship between human likeness and familiarity (Fig. 1). Although Mori proposed that abstract human forms should serve as the principle for designing socially-acceptable robots, the uncanny valley can be seen more positively—for example, by indicating when a robot's responses do not rise to the expectations elicited by its human form. This provides useful feedback for improving the cognitive models implemented in the android (see Fig. 6).

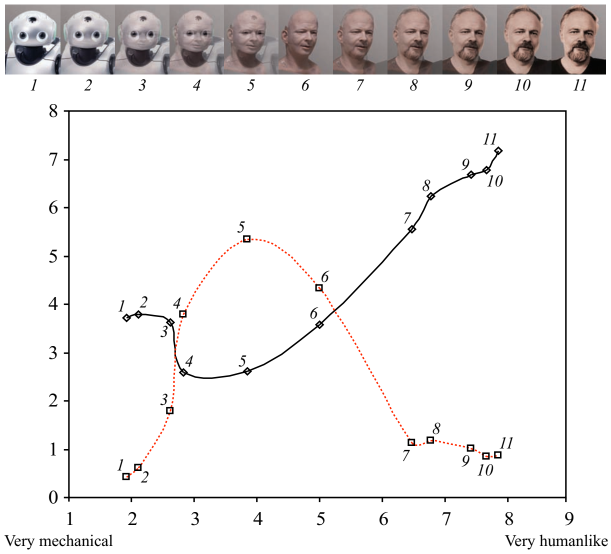

Figure 2: Average ratings for *strange versus familiar* (solid line) and *eeriness* (dashed line) were plotted against *mechanical versus humanlike* for images morphing from the robot Qrio the Philip K. Dick android to Philip K. Dick himself. The plots reproduce Mori's hypothesized uncanny valley and indicate a corresponding region of eeriness.

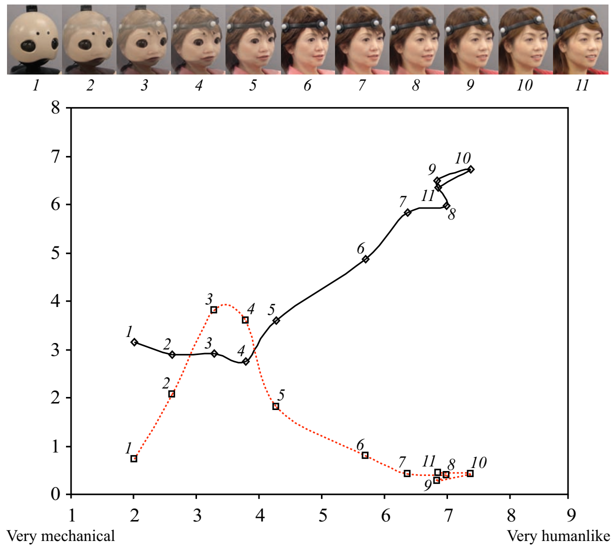

Figure 3: The same plots as Fig. 2 but for images morphing from the robot Eleviee to the android Repliee Q1Expo to the android's human model.

Plotting the uncanny valley

In a previous study, 45 Indonesian participants were asked to rate 31 images on a nine-point scale ranging from very mechanical (1) to very humanlike (9) and from very strange (1) to very familiar (9) [MacDorman and Ishiguro, 2006]. They were then asked to select eerie images and rate them for eeriness on a ten-point scale, ranging from slightly eerie (1) to extremely eerie (10).[1] Two sets of 11 morphed images were included among the 31 images.

Fig. 2 shows the plot of the average ratings on the *strange versus familiar* (solid line) and *eeriness* (dashed line) scale for the first set of images, which morphed from a photograph of the humanoid robot Qrio (left) to one of the Philip K. Dick android developed by Hanson Robotics (center) to one of Philip K. Dick himself (right). Fig. 3 shows the plot for a second set of images, which morphed from a photograph of the humanoid robot Eveliee (left) to one of the android Repliee Q1Expo (center) to Repliee's human model (right). The independent axis is the average rating on the *mechanical versus humanlike* scale. The plots reproduce Mori's posited uncanny valley (solid line) and indicate a region of eeriness in the same area (dashed line).

Experiment: Ratings of videos

The intention of the current study is to determine whether the uncanny valley is a necessary property of near-humanlike forms. Participants are presented with short video clips of a wide range of mainly android and humanoid robots engaged in various activities in different settings.

[1] Images that were not selected as eerie were rated 0.

Subjects. There were 56 Indonesian participants, 43 male and 13 female, of whom 13 were 17 to 20 years old, 36 were 21 to 25, 4 were 26 to 30, and 3 were 31 to 35. The participants were mainly university students, young professionals, and government workers. Participants were recruited from an Internet cafe and received two hours of free Internet access.

Procedure. Participants were asked on a computer-based questionnaire, in individual sessions, to rate 14 video clips, most of which were 30 to 60 seconds in length, on a nine-point *mechanical versus humanlike* scale, a nine-point *strange versus familiar* scale, and a ten-point *eeriness* scale. The scales ranged from very mechanical (1) to very humanlike (9), from very strange (1) to very familiar (9), and from not eerie (0) to extremely eerie (10). The video clips included a mobile robot (Pioneer II), a manipulator arm, seven humanoid robots (Rovovie-M3, HR-2, VisiON Nexta, Chronio, Robovie, Wakamaru, Asimo), two android heads (K-bot, Eva), two androids (Philip K. Dick, Repliee Q1Expo), and one human being. The video clips were presented in random order. For each video the three ratings were requested in random order. The direction of the scales was determined randomly for each question.

Results. Fig. 4 shows the plot of the average ratings on the *strange versus familiar* (solid line) and *eeriness* (dashed line) scale for a given average rating on the *mechanical versus humanlike* scale for the video clips of the 14 robots. There is no consistent valley shape when plotting familiarity against human likeness. Instead, there are oscillations in eeriness for robots that range from mechanical looking (1.96) to approaching very humanlike (8.57). The plot of *strange versus familiar* and *eeriness* are almost mirror images.

The average standard deviation (SD) was 1.69 for the *mechanical versus humanlike* scale, 2.43 for the *strange*

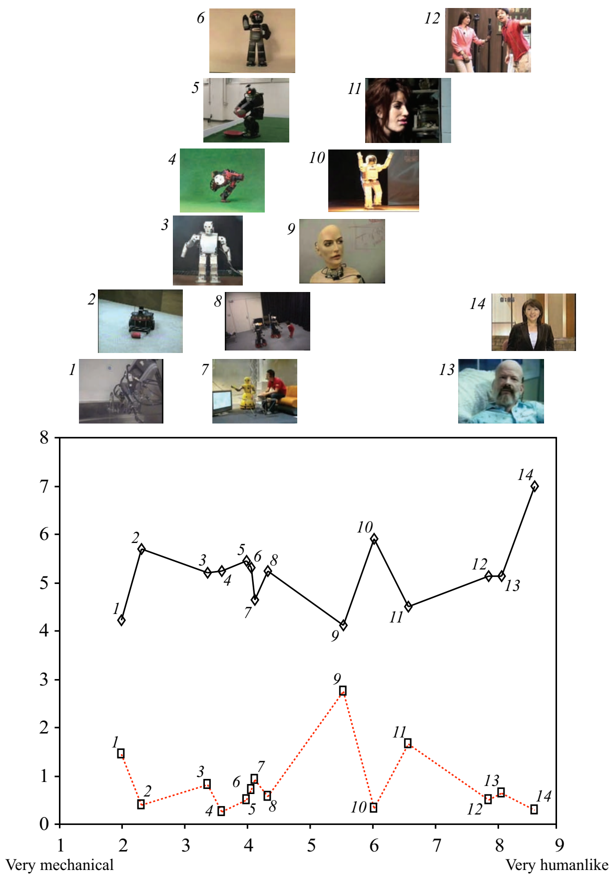

Figure 4: Based on the average ratings of 56 participants, 14 video clips are arranged from left, mechanical, to right, approaching very humanlike. The names of the robots are listed in Table 1. The solid line plots the relationship between perceived humanlikeness (on the *mechanical versus humanlike* scale) and perceived familiarity (on the *strange versus familiar* scale). The dashed line plots the relationship between perceived humanlikeness and eeriness. There is no single uncanny valley in the plot.

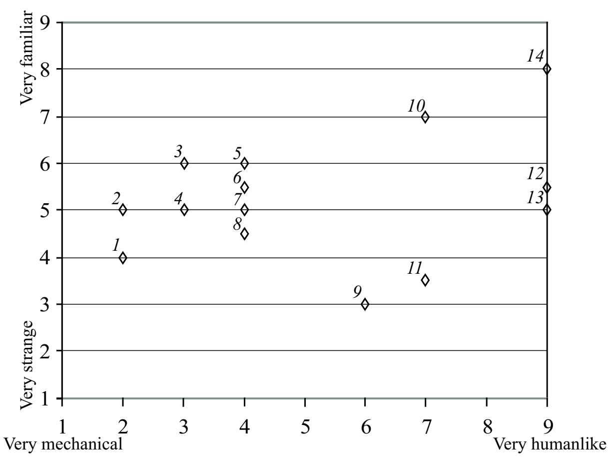

Figure 5: The median ratings on a *strange versus familiar* scale are plotted against the median ratings on a *mechanical versus humanlike* scale for the robots in Table 1 and Fig. 4. Robots with the same median value for human likeness can show quite different median values for familiarity.

Table 1: Median ratings of the 14 video clips on a very mechanical (1) to very humanlike (9) scale and a very strange (1) to very familiar (9) scale

No.	Name	mech. v hum.	strange v fam.
1	Manipulator	2	5
2	Pioneer II	2	4
3	HR-2	3	6
4	Robovie-M3	3	5
5	Nexta	4	6
6	Chronio	4	5.5
7	Wakamaru	4	5
8	Robovie	4	4.5
9	K-bot	6	3
10	Asimo	7	7
11	Eva	7	3.5
14	Human	9	8
12	Repliee	9	5.5
13	PKD android	9	5

versus familiar scale, and 1.68 for the *eeriness* scale.

Median values were considered a more robust indicator of central tendency, especially given the high variance in the data and the the subjective nature of the questions. Table 1 lists the values for the *mechanical versus humanlike* and *strange versus familiar* scales. The median values for the *eerie* scale are not listed because they were all 0 except for K-bot whose median value was 1. Fig. 5 plots the median values for the *strange verus familiar* scale against the median values for the *mechanical versus humanlike* scale.

Table 1 and Fig. 5 show that video clips of robots that were rated as having the same median human likeness could have much more or much less median familiarity. For example, the median value was *very humanlike* (9) for the Philip K. Dick android, Repliee Q1Expo, and its human model, although the median value for the human model was *familiar* (8) but *neutral* (5) for the Philip K. Dick android and near neutral (5.5) for Repliee Q1Expo. The median values also represent Asimo as *somewhat familiar* and K-bot as *somewhat strange*, although Asimo was represented as more humanlike than K-bot. Thus, the depiction of Asimo in the video clip seemed to bump up its human likeness despite the fact that it does not have a humanlike face with skin, teeth, nostrils, pupils, and so on.

Discussion

The video clips exhibit a wide range of robots performing different actions in quite different contexts, sometimes with speech accompaniment. The results do not indicate a single uncanny valley for a particular range of human likeness. Rather, they suggest that human likeness is only one of perhaps many factors influencing the extent to which a robot is perceived as being strange, familiar, or eerie. This is an important result because it implies that factors other than human likeness could be

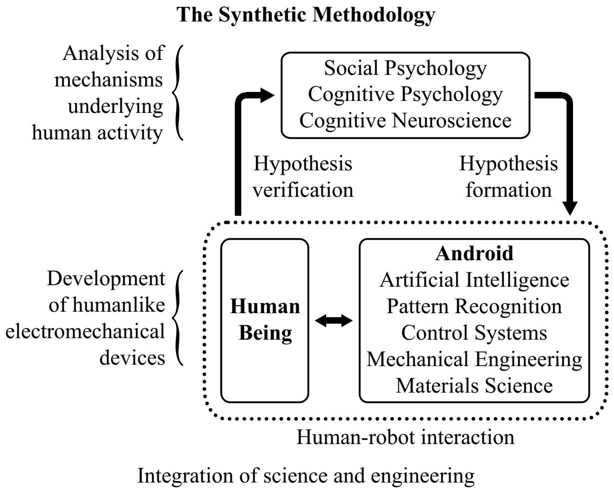

Figure 6: Ishiguro [2005] proposes a synthetic methodology for investigating human interaction that integrates science and engineering.

manipulated to overcome the uncanny valley. Of course, Mori [1970] [Mori, 1970] identified motion as one such factor, so it may be argued that his hypothesis could accommodate the results. But the variations in familiarity and eeriness for a given level of human likeness are not consistent with motion-induced magnifications of a valley shape.

So why does morphing from a mechanical-looking robot to an android and then to its human model produce an uncanny valley in still images, as shown in Fig. 2 and 3? One possible explanation is that the morphs provide a more gradual and consistent change with less extraneous variation. For example, as the images were still, variations in movement and speech were not an issue. We might expect to find uncanny valleys in more controlled experiments that vary appearance or movement along fewer parameters.

The notion that the uncanny valley can be escaped by varying factors unrelated to human likeness is consistent with an experiment performed by Hanson [2006] using morphs. Although he found that morphing from a mechanical-looking robot to an android produced a valley in a familiarity scale and in an appealing scale and a peak in an eeriness scale, these effects were greatly reduced by tuning the morphs. Thus, without making a morph more or less humanlike, Hanson was largely able to design around the uncanny valley. His technique was to adjust the appearance of the uncanny morphs toward the cuter features of a doll.

Conclusion

The results of the experiment suggest that human likeness is only one factor determining the familiarity, strangeness, and eeriness of a robot. This offers the hope that other factors modulating these qualities will be uncovered. MacDorman and Ishiguro [2006] have documented a number of possible explanations for the uncanny valley, ranging from expectation violation and cognitive paradoxes [Ramey, 2005] to evolutionary aesthetics [Etcoff, 1999] and pathogen avoidance. As the validity of these explanations comes under scientific scrutiny, design principles will appear that engineers can use to develop robots with desirable aesthetics. Whether it is indeed desirable to build a robot that is appealing or unnerving will depend on its purpose and the setting in which it is used. Given the success of the horror genre, it is clear that eeriness is not always considered a bad thing.

Acknowledgments

Much appreciation goes to Heryati Madiapuri for recruiting participants and conducting the experiments in September, 2005, to Yuli Suliswidiawati, for reviewing the experimental procedures, and to Z. A. Dwi Pramono and Christopher H. Ramey for serving on the ethics advisory panel and providing comments on the experimental design. Thanks also go to David Hanson, Almir Heralik, Takayuki Kanda, Hiroshi Ishiguro, Takahashi Tomotaka, ActivMedia Robots, ATR, Honda, Kokoro Co., Ltd., Mitsubishi, and Vstone for the robots.

References

[Cowley and MacDorman, 2006] Cowley, S. and MacDorman, K. (2006). What baboons, babies, and tetris players tell us about interaction: A biosocial view of norm-based social learning. *Connection Science*.

[Etcoff, 1999] Etcoff, N. L. (1999). *Survival of the prettiest: The science of beauty*. Doubleday, New York.

[Hanson, 2006] Hanson, D. (2006). Exploring the aesthetic range for humanoid robots. In *Proceedings of the ICCS/CogSci-2006 Symposium: Toward Social Mechanisms of Android Science*, Vancouver, Canada.

[Ishiguro, 2005] Ishiguro, H. (2005). Android science: Toward a new cross-disciplinary framework. In *CogSci-2005 Workshop: Toward Social Mechanisms of Android Science*, pages 1–6, Stresa, Italy.

[MacDorman and Ishiguro, 2006] MacDorman, K. and Ishiguro, H. (2006). The uncanny advantage of using androids in social and cognitive science research. *Interaction Studies*, 7.

[MacDorman et al., 2005] MacDorman, K., Minato, T., Shimada, M., Itakura, S., Cowley, S., and Ishiguro, H. (2005). Assessing human likeness by eye contact in an android testbed. In *Proceedings of the XXVII Annual Meeting of the Cognitive Science Society*, Stresa, Italy.

[Mori, 1970] Mori, M. (1970). Bukimi no tani [the uncanny valley]. *Energy*, 7:33–35.

[Ramey, 2005] Ramey, C. H. (2005). The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. In *IEEE-RAS International Conference on Humanoid Robots*, Tsukuba, Japan.