A challenge to the study of individual differences in uncanny valley sensitivity: The importance of looking at individual-level response patterns

Tyler J. Burleigh
University of Guelph, Ontario, Canada

Keywords: uncanny valley; individual differences; sensitivity; human-likeness; affect; categorization

The uncanny valley refers to a nonlinear affective response function that is observed across a range of stimuli that vary in human-likeness or degree of category membership. The "valley" itself refers to an intermediate region along this response function wherein stimuli receive the most negative affect. This phenomenon has generated interest among researchers and laypeople alike, which I suspect is partly because it is a counter-intuitive finding (i.e., why isn't human-likeness monotonically related to positive affect?), and partly because it has been useful as a way to describe certain affective experiences (e.g., individuals frequently use "uncanny" or "eerie" as adjectives to describe their sense that something is "not quite right" about certain robotic or computer-animated entities).

Although the phenomenon was originally described anecdotally by Masahiro Mori in 1970, it seems that serious inquiry began only after Karl MacDorman translated Mori's writing into English (Appendix B in MacDorman, 2005) and published his own experimental research that provided preliminary evidence for the phenomenon's existence and also investigated several of its plausible causes (e.g., MacDorman & Ishiguro, 2006; MacDorman, Green, Ho, & Koch, 2009). While research into the uncanny valley phenomenon has seen great progress, there are many areas of inquiry that have yet to be explored. In the target article, Karl MacDorman and Steven Entezari (this volume) examine a largely neglected aspect of the phenomenon: individual differences in sensitivity to uncanny stimuli.

Existing research on the phenomenon has almost exclusively been carried out using experimental and quasi-experimental methods, such as by generating or selecting stimuli that vary in human-likeness or categorical membership, and then measuring individuals' cognitive and affective responses to them (e.g., Burleigh, Schoenherr, & Lacroix, 2013; Burleigh & Schoenherr, 2014). In context of this work, researchers have placed an emphasis on the mean affective responses that are observed across the stimulus continua, and individual differences are treated as noise around the central tendency of stimulus-response distributions.

In contrast to this experimental work, MacDorman and Entezari (this volume) adopted an individual differences approach, which treats the variation in individual-level responding as a meaningful object of analysis. In addition to measuring human-likeness judgments and affective responses to robot, android, and human stimuli, the authors also measured nine individual traits. These included personality dimensions (neuroticism, perfectionism, and anxiety), worldviews and beliefs (religious fundamentalism and human-robot / human-android uniqueness), attitudes towards robots, and other individual sensitivities (personal distress and sensitivity to reminders of one's "creatureliness"). The authors provided reasonable justifications for their choice of individual difference measures, and the data supported several of their hypotheses. However, the authors decided to exclude responses to nonandroid stimuli from their analyses of individual differences. I believe this decision was problematic, as I will explain in the following section.

Do Individuals' Response Patterns Reflect Mean Response Patterns?

As I mentioned previously, the phenomenon itself is defined as a nonlinear affective response function that is observed across a range of stimuli that vary in human-likeness or degree of category membership. This convention is due to the history of the concept, which originates from the function that was articulated by Mori (1970). One of the primary sources of evidence for the phenomenon has therefore been the observation that patterns of affective responses reflect the characteristics of Mori's (1970) nonlinear pattern. However, researchers have mainly analyzed *mean* response patterns, assuming that they are an accurate proxy for the response patterns of individuals (for an exception to this, see Ferrey, Burleigh, & Fenske, 2014). This is perhaps a shortcoming of uncanny valley research in general, but it is a serious problem for research on individual differences in uncanny valley sensitivity.

Along these lines, I would question MacDorman and Entezari's (this volume) decision to exclude nonandroid stimuli from their analyses of individual differences. Originally, the authors had selected six video clips (two robots, three androids, and one human) to represent a continuum of human-likeness. Participants watched these video clips, and rated their eeriness, warmth, and human-likeness. It was observed that the patterns of mean responses (see Figure 2, MacDorman & Entezari, this volume) were consistent with an uncanny valley interpretation. For example, the android stimuli received, on average, higher human-likeness and eeriness ratings than the robot stimuli but, on average, lower human-likeness and eeriness ratings than the human stimulus. The authors then excluded nonandroid stimuli from subsequent analyses, arguing that this decision "was made on theoretical grounds, namely, because they are not depicted in the uncanny valley" and further state that it "was supported by the data" (p. 19).

I do not believe that their data supports the decision to exclude nonandroid stimuli. Such a decision assumes that mean response patterns reflect individual response patterns. Yet, the mean response patterns could have been an artifact of averaging across individuals. It is possible that many individuals' response patterns would not support an uncanny valley interpretation, even those who rated android stimuli as high on eeriness, because, for example, those same individuals could have also rated the robot stimuli as high on eeriness. I would argue that subjective responses to android stimuli alone are not an adequate basis for operationalizing “sensitivity to the uncanny valley”.

Hypothetical Data Set: An Illustration

As an illustration of my concerns, consider the following hypothetical data set, and the mean response plot and individual response plots that are generated from it.

Table 1. Hypothetical data; individual eeriness and human-likeness ratings for each stimulus.

		Participants				
Stimuli	Ratings	P1	P2	P3	P4	P5
S1 (robot)	Eeriness	3	1	2	3	2
	Human-Likeness	1	2	1	1	1
S2 (android)	Eeriness	7	4	1	4	4
	Human-Likeness	1	5	7	4	4
S3 (human)	Eeriness	1	1	1	1	4
	Human-Likeness	7	7	7	7	7

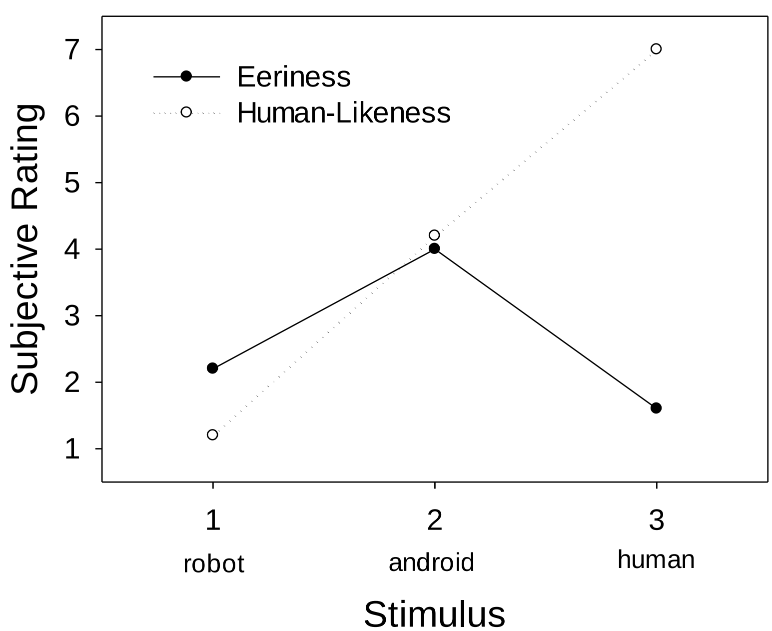

Figure 1. Mean plot generated from hypothetical data, indicating response patterns that are consistent with an uncanny valley interpretation.

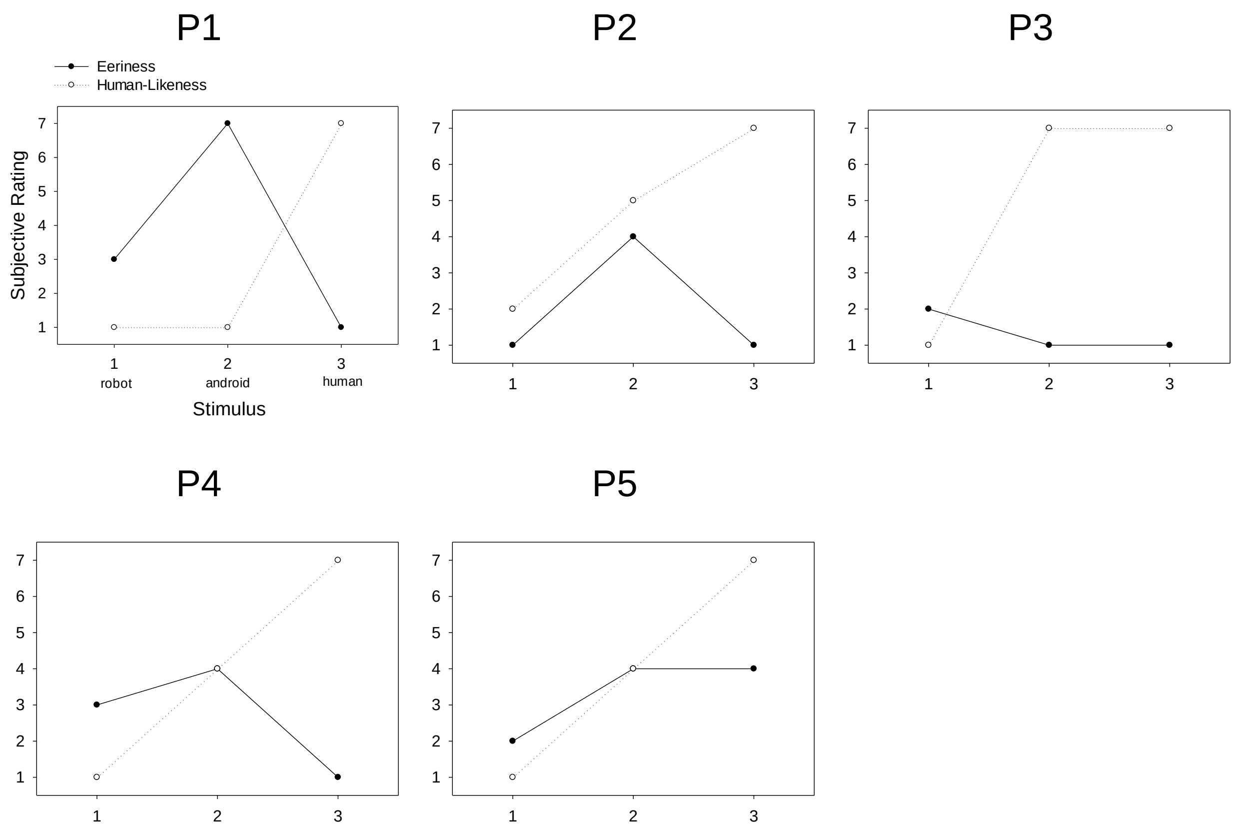

Figure 2. Plots for individual participants generated from hypothetical data. Response patterns for P2 and P4 are consistent with an uncanny valley interpretation, but response patterns for P1, P3, and P5 are not.

In this hypothetical example, it can be seen that the mean response plot is consistent with an uncanny valley interpretation (see Figure 1). Namely, stimuli at intermediate regions along the stimulus continuum (S3; the “android”) were, on average, associated with both greater human-likeness and greater eeriness than the stimulus at the left anchor (S1; the “robot”). Further, the stimulus at the right anchor (S5; the “human”) was, on average, associated with the greatest human-likeness and the least eeriness. However, if this mean response plot is decomposed into response plots for each of the individuals (see Figure 2), then it becomes clear that the mean response pattern does not accurately reflect the response patterns of individuals. Indeed, in this example, only two out of five (40%) of the hypothetical participants (i.e., P2 and P4) observed a pattern that would be consistent with an uncanny valley interpretation. The remaining three participants (i.e., P1, P3, and P5) did not observe an uncanny valley. For example, although P1 showed the expected pattern for eeriness, they also rated the android as equally human-like as the robot; P5 showed the expected pattern for human-likeness, but also rated the android and human stimuli as equally eerie; and P3 did not show the expected pattern for either eeriness or human-likeness.

It stands to reason that similar misclassifications could have been made in MacDorman and Entezari's (this volume) study. In their study, nonandroid stimuli were used as a point of comparison for determining that android stimuli were, on average, the "uncanny" stimuli. But rather than classify the *stimuli* in terms of their "uncanny" status, and then proceeding to examine responses to them, I would argue that the correct approach to determining if individuals are "sensitive to the uncanny valley" would be to classify *individuals* based on their response patterns across the range of stimuli, and then examining the correlates of such classifications.

Individual sensitivity to the uncanny valley could be operationalized as the presence or absence of an uncanny valley pattern, a discrete variable. In my own research program, I have preferred curve-fitting as a formal method of testing for nonlinear patterns (Burleigh & Schoenherr, 2014; Burleigh et al., 2013; Ferrey et al., 2014). Developing a continuous variable would then require some creativity. For example, it might be reasonable to first classify individuals whose responses reflect the uncanny valley pattern, and then for those individuals calculate the mean difference between their affective ratings of android and robot stimuli. In the hypothetical data set, for example, it can be seen that while P2 and P4 each display the uncanny valley pattern, and while they each provide the same ratings of android stimuli, the difference between P2's rating of android and robot stimuli was greater than P4's. This type of calculation would allow researchers to examine individuals' sensitivity to the uncanny valley at a finer level of analysis.

Conclusions

In conclusion, I believe that the complexity that is inherent to the definition of the uncanny valley phenomenon poses a challenge to the study of individual differences in uncanny valley sensitivity. It is not enough to examine variation in responses to android stimuli, as MacDorman and Entezari (this volume) have done in the target article, because such variation is only meaningful in relation to individuals' responses to stimuli at other levels of human-likeness or category membership. I have proposed a solution to this problem which would involve classifying individuals based on individual response patterns, and then potentially calculating continuous measures of sensitivity that account for the relative differences in individuals' ratings of stimuli. Thus, although this is a challenge for researchers to address in future studies of uncanny valley sensitivity, it is a challenge that can be overcome.

References

Burleigh, T. J., & Schoenherr, J. R. (2015). A reappraisal of the uncanny valley: categorical perception or frequency-based sensitization? *Frontiers in Psychology*, 5:1488. doi: 10.3389/fpsyg.2014.01488

Burleigh, T. J., Schoenherr, J. R., & Lacroix, G. L. (2013). Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behavior*, *29*(3), 759–771. doi: 10.1016/j.chb.2012.11.021

Ferrey, A., Burleigh, T. J., & Fenske, M. (2015). Stimulus-category competition, inhibition, and affective devaluation: a novel account of the uncanny valley. *Frontiers in Psychology,* 6:249. doi: 10.3389/fpsyg.2015.00249

MacDorman, K. F. (2005, July). Androids as an experimental apparatus: Why is there an uncanny valley and can we exploit it? In *CogSci-2005 workshop: toward social mechanisms of android science* (pp. 106-118).

MacDorman, K. F., & Entezari, S. O. (this volume). Individual differences predict sensitivity to the uncanny valley, *Interaction Studies*.

MacDorman, K. F., Green, R. D., Ho, C.-C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior*, *25*(3), 695–710. doi: 10.1016/j.chb.2008.12.026

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, *7*(3), 297–337. doi: 10.1075/is.7.3.03mac

Mori. (1970). Bukimi no tani [The uncanny valley]. *Energy*, *7*(4), 33–35.

Author's address:

Tyler J. Burleigh
Department of Psychology
University of Guelph
50 Stone Road East
Guelph, ON N1G 2W1

E-mail: tburleig@uoguelph.ca

Biographical note:

Tyler J. Burleigh received his Master's in Cognitive Psychology from Carleton University in 2011. His thesis empirically examined the hypothesis that was put forward by Masahiro Mori in 1970 to explain the uncanny valley phenomenon based on human-likeness, as well as an alternative hypothesis based on category conflict. Currently, he is a PhD candidate at the University of Guelph where he studies the uncanny valley, social cognition, and environmental psychology.