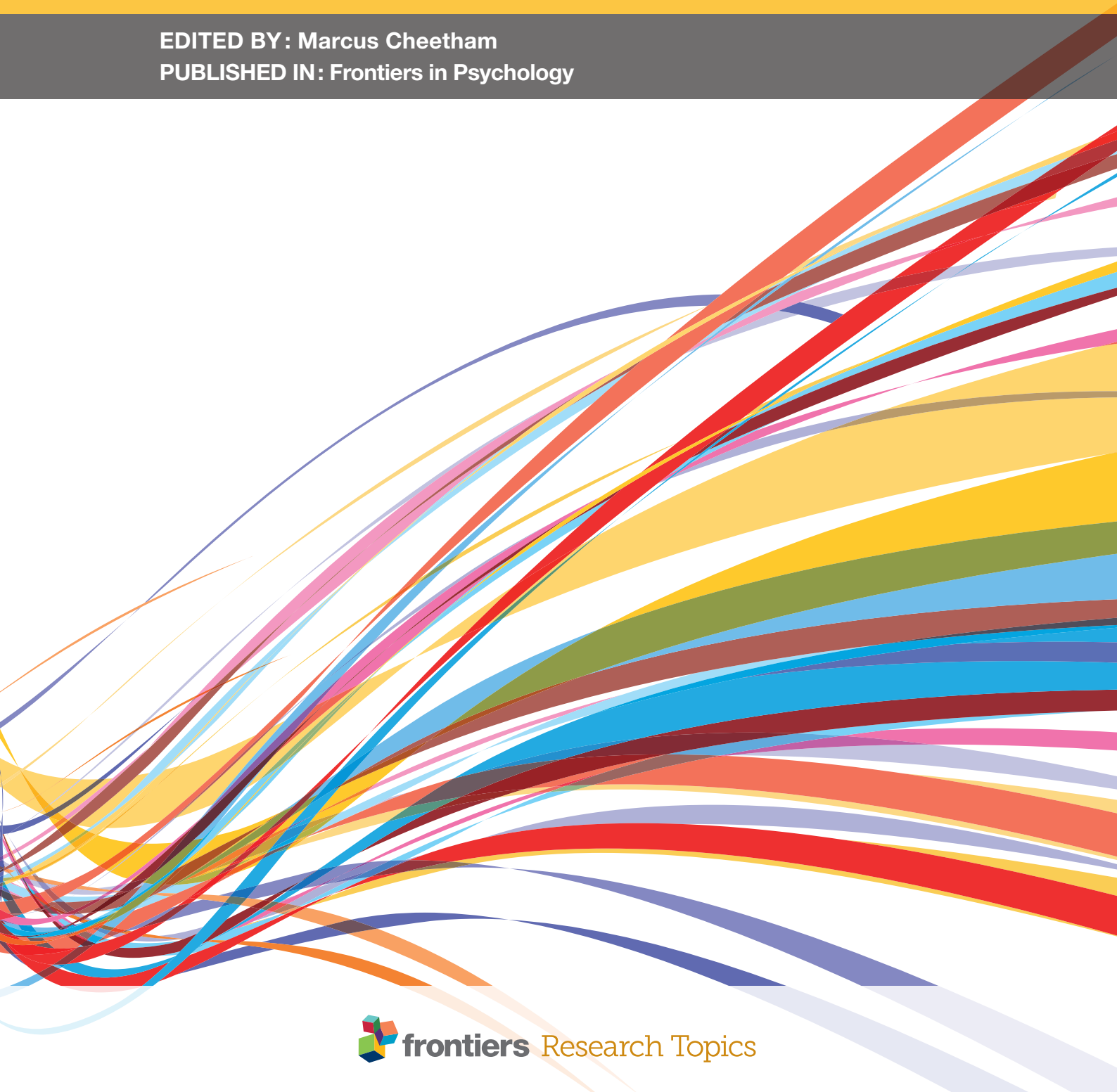
THE UNCANNY VALLEY HYPOTHESIS AND BEYOND

EDITED BY: Marcus Cheetham
PUBLISHED IN: Frontiers in Psychology



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88945-443-3
DOI 10.3389/978-2-88945-443-3

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: **researchtopics@frontiersin.org**

THE UNCANNY VALLEY HYPOTHESIS AND BEYOND

Topic Editor:
Marcus Cheetham, University Hospital Zurich and University of Zurich, Switzerland

A field of theory and research is evolving around the question highlighted in the Uncanny Valley Hypothesis: How does high realism in anthropomorphic design influence human experience and behaviour? The Uncanny Valley Hypothesis posits that a very humanlike character or object (e.g., robot, prosthetic limb, doll) can evoke a negative affective (i.e., uncanny) state. Recent advances in robotic and computer-graphic technologies in simulating aspects of human appearance, behaviour and interaction have been accompanied, therefore, by theorising and research on the meaning and relevance of the Uncanny Valley Hypothesis for anthropomorphic design. Current understanding of the "uncanny" idea is still fragmentary and further original research is needed. However, the emerging picture indicates that the relationship between humanlike realism and subjective experience and behaviour may not be as straightforward as the Uncanny Valley Hypothesis suggests. This Research Topic brings together researchers from traditionally separate domains (including robotics, computer graphics, cognitive science, psychology and neuroscience) to provide a snapshot of current work in this field. A diversity of issues and questions are addressed in contributions that include original research, review, theory, and opinion papers.

Citation: Cheetham M., ed. (2018). The Uncanny Valley Hypothesis and Beyond. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-443-3

Table of Contents

frontiers in Psychology

EDITORIAL
published: 17 October 2017
doi: 10.3389/fpsyg.2017.01738

Editorial: The Uncanny Valley Hypothesis and beyond

Marcus Cheetham [1, 2, 3]*

[1] Department of Internal Medicine, University Hospital Zurich, Zurich, Switzerland, [2] University Research Priority Program Dynamics of Healthy Aging, University of Zurich, Zurich, Switzerland, [3] Department of Neuropsychology, Institute of Psychology, University of Zurich, Zurich, Switzerland

Keywords: uncanny valley hypothesis, robotics, computer animation, computer graphics, virtual reality, human likeness, anthropomorphic design

Editorial on the Research Topic

The Uncanny Valley Hypothesis and beyond

Progress toward realistic simulation of human appearance, behavior, and interaction in the fields of robotics and computer-graphics has been accompanied by interest in the *Uncanny Valley Hypothesis* (*UVH*) (Mori, 1970). The UVH posits that the use of anthropomorphic realism in the design of characters and objects (e.g., robots, prostheses) might have a counterproductive effect. Instead of enhancing subjective experience of the character or object, certain degrees of greater realism might unsettle the observer and induce a negative affective state. This state is marked by feelings of personal disquiet and a sense of strangeness (i.e., an uncanny effect). Mori did not develop the UVH further or subject his idea to empirical test. But concern as to the UVH' potential relevance for anthropomorphic design has given impetus to a new and evolving field of research (e.g., Hanson, 2006; MacDorman, 2006).

The UVH describes affective experience and the relationship between this and humanlike realism in simple terms. The simplicity serves well to express the general notion of a potential problem in anthropomorphic design. But the UVH is not a hypothesis in the scientific sense of an empirically testable statement. Early scientific enquiry into the UVH placed emphasis on developing experimentally more tangible renderings of the UVH. This can be seen in early exploratory testing, debate about the meaning and measurement of the UVH′s affective dimension, and theoretical considerations as to mechanisms potentially conducive to uncanny experience (e.g., Hanson, 2006; MacDorman, 2006; Bartneck et al., 2007; Seyama and Nagayama, 2007). In this research topic, the review, opinion, hypothesis and theory, and original research articles make reference to the early work.

Different lines of empirical enquiry into anthropomorphic realism and human behavior have built on the early work. These include the behavioral and physiological investigation of a range of perceptual, cognitive and affective mechanisms (e.g., Chaminade et al., 2007; Looser and Wheatley, 2010; Saygin et al., 2012). In this research topic, de Borst and de Gelder review behavioral and physiological evidence of differences and similarities in the response of subjects to the appearance and the affective and motor behavior of human-like compared with natural human stimuli. Understanding the pattern of responses to these perceptual categories is important. Much work has focussed on the affective experience of more or less realistic nonhuman stimuli (e.g., Tinwell and Grimshaw, 2009). However, the perception and experience of humanlike realism and related affect within the human category itself has received very little attention (Cheetham et al., 2011), even though this category forms the point of reference for the UVH and the proposed problem in anthropomorphic design.

OPEN ACCESS

Edited and reviewed by:
Eddy J. Davelaar, Birkbeck University of London, United Kingdom

***Correspondence:**
Marcus Cheetham
m.cheetham@psychologie.uzh.ch

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *21 March 2017*
Accepted: *20 September 2017*
Published: *17 October 2017*

Citation:
Cheetham M (2017) Editorial: The Uncanny Valley Hypothesis and beyond. Front. Psychol. 8:1738. doi: 10.3389/fpsyg.2017.01738

Empirical support for the uncanny idea has been inconsistent. A number of specific reasons for this are addressed in different articles of this topic. Generally, some inconsistency will be inherent to the use of different conceptual, technical, and methodological approaches to define and operationalize humanlike realism and human experience and behavior. But there appears to be a growing convergence toward the observation, description, and explanation of the uncanny idea in terms of properties and mechanisms of perceptual and category information processing. This is reflected in the review article by Kätsyri et al. in this research topic. Kätsyri et al. consider inconsistent findings, review different conceptualisations of the uncanny effect, and report whether these conceptualisations find empirical support.

One particular line of enquiry, with a relatively brief history of investigation in the field of the UVH, has focussed on what might be generally referred to as the *categorization difficulty hypothesis*. This posits that subjective difficulty assigning a categorically ambiguous stimulus to the human or non-human category induces a negative affective state. Typically, this hypothesis has been examined using stimuli to represent the nonhuman and human categories of the UVH' dimension of human likeness (e.g., Burleigh et al., 2013). In their hypothesis and theory article, Ferrey et al. explore whether the experience of negative affect relates to cognitive mechanisms that subserve the processing of conflicting information from different perceptual categories irrespective of whether these categories contain perceptual features that specify human likeness. To date, the categorization difficulty hypothesis has been considered from a cognitive perspective in relation to various concepts, such as processing fluency (e.g., Yamada et al., 2013). In their opinion article, Schoenherr and Burleigh present a social-cultural perspective on the occurrence of negative affect in relation to low previous exposure to categorically ambiguous stimuli (i.e., an inverse mere-exposure effect).

Original research in the field of the UVH has made wide use of *ad-hoc* developed self-rating scales (e.g., Looser and Wheatley, 2010). With little exception (e.g., Ho and MacDorman, 2010), no psychometrically valid and reliable measures of affect have been applied in uncanny-related research. In original research of this topic, Cheetham et al. use well-established physiological and validated behavioral measures to examine affect in relation to the processing of nonhuman and human perceptual category information and of conflicting category information. To overcome the interpretational issues that have dogged the conceptualization of uncanny feelings in the UVH, affect is examined in terms of the primary orthogonal dimensions of affective experience (Russell, 1980).

While a typical human observer has everyday expertise in the extraction, processing, and interpretation of human perceptual and category information, the observer has comparably little or no such experience in the processing of newly designed humanlike exemplars and their human-specifying perceptual cues. This asymmetry between the nonhuman and human in perceptual and categorisation experience and knowledge and its influence on affective experience has received scant attention to date (Cheetham et al., 2011). In this topic, Cheetham et al. apply a perceptual discrimination paradigm to investigate this asymmetry in terms of differences within and between non-human and human perceptual categories in perceptual sensitivity to human-specifying information and examine the relationship between this and affect. Burleigh and Schoenherr use a category-learning paradigm to examine the modulatory influence on negative affect of acquired category structure (using perceptual categories based on non-human stimuli) and repeated stimulus exposure. Similarly, Złotowski et al. investigate the influence of repeated exposure to human-robot interaction on affective experience, in their case applying a live interaction paradigm. Generally, these studies focus on psychological factors that influence subjective experience of nonhuman and human stimuli. In contrast, Handžić and Reed focus on the systematic variation of multiple parameters of normal and abnormal patterns of gait and show how varying these parameters may be used to influence subjective experience. Finally, Destephe et al. study the impact of a range of factors, such as intensity of emotion expressed by whole-body robotic movement, in relation to different categories of subjective experience and affective determinants (e.g., attitudes).

The contributions to this topic provide a snapshot of current research in the field of the UVH. Set in the context of previous work, this snapshot suggests at least one emerging trend in research. Work to ascertain the general relevance of the uncanny idea for anthropomorphic design dominated the early years of uncanny research. A comparative interpretation of findings to date and inconsistencies between these suggests that an uncanny effect is not generalizable across different individuals, stimuli, situations, tasks, and time. As this topic indicates, research is shifting toward the development of a differentiated understanding of specifically when, under what conditions and why effects consistent with the uncanny idea occur. The development of this understanding requires research and active publication of a broad range of findings. These include robust findings that support the uncanny idea, indicate no effects, show effects contrary to the uncanny idea, and findings that do not replicate previously published findings. By broadening the perspective beyond the focus on negative affect, in the narrow sense of the UVH, this informative approach can contribute to the development of knowledge about anthropomorphic effects that are consistent or at variance with the uncanny idea, promote therefore a differentiated understanding of the relevance of the uncanny idea for anthropomorphic design, and help to accumulate applied knowledge about the effective use of variously realistic anthropomorphic features to induce, direct, maintain, and motivate subjective experience and behavior.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

REFERENCES

Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. (2007). "Is the uncanny valley an uncanny cliff?," in *Proceedings of the 16th IEEE, RO-MAN (Jeju),* 368–373.

Burleigh, T. J., Schoenherr, J. R., and 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. *Comput. Hum. Behav.* 29, 759–771. doi: 10.1016/j.chb.2012.11.021

Chaminade, T., Hodgins, J., and Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters' actions. *Soc. Cogn. Affect. Neurosci.* 2, 206–216. doi: 10.1093/scan/nsm017

Cheetham, M., Suter, P., and Jancke, L. (2011). The human likeness dimension of the "uncanny valley hypothesis": behavioral and functional MRI findings. *Front. Hum. Neurosci.* 5:126. doi: 10.3389/fnhum.2011.00126

Hanson, D. (2006). "Exploring the aesthetic range for humanoid robots," in *Paper Presented at the Proceedings of the ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science* (Vancouver, BC).

Ho, C.-C., and MacDorman, K. (2010). Revisiting the uncanny valley theory: developing and validating an alternative to the Godspeed indices. *Comput. Hum. Behav.* 26, 1508–1518. doi: 10.1016/j.chb.2010.05.015

Looser, C. E., and Wheatley, T. (2010). The tipping point of animacy: how, when, and where we perceive life in a face. *Psychol. Sci.* 21, 1854–1862. doi: 10.1177/0956797610388044

MacDorman, K. (2006). "Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: an exploration of the uncanny valley," in *Paper Presented at the ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science* (Vancouver, BC).

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

Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178.

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., and Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* 7, 413–422. doi: 10.1093/scan/nsr025

Seyama, J., and Nagayama, R. S. (2007). The uncanny valley: the effect of realism on the impression of artificial human faces. *Presence* 16, 337–351. doi: 10.1162/pres.16.4.337

Tinwell, A., and Grimshaw, M. (2009). "Bridging the uncanny: an impossible traverse?" in *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era*, eds O. Sotamaa, A. Lugmayr, H. Franssila, P. Näränen, and J. Vanhala (Tampere: ACM), 66–73.

Yamada, Y., Kawabe, T., and Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *Jpn. Psychol. Res.* 55, 20–32. doi: 10.1111/j.1468-5884.2012.00538.x

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

REVIEW
published: 12 May 2015
doi: 10.3389/fpsyg.2015.00576

Is it the real deal? Perception of virtual characters versus humans: an affective cognitive neuroscience perspective

Aline W. de Borst and Beatrice de Gelder*

Brain and Emotion Laboratory, Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, Netherlands

Recent developments in neuroimaging research support the increased use of naturalistic stimulus material such as film, avatars, or androids. These stimuli allow for a better understanding of how the brain processes information in complex situations while maintaining experimental control. While avatars and androids are well suited to study human cognition, they should not be equated to human stimuli. For example, the uncanny valley hypothesis theorizes that artificial agents with high human-likeness may evoke feelings of eeriness in the human observer. Here we review if, when, and how the perception of human-like avatars and androids differs from the perception of humans and consider how this influences their utilization as stimulus material in social and affective neuroimaging studies. First, we discuss how the appearance of virtual characters affects perception. When stimuli are morphed across categories from non-human to human, the most ambiguous stimuli, rather than the most human-like stimuli, show prolonged classification times and increased eeriness. Human-like to human stimuli show a positive linear relationship with familiarity. Secondly, we show that expressions of emotions in human-like avatars can be perceived similarly to human emotions, with corresponding behavioral, physiological and neuronal activations, with exception of physical dissimilarities. Subsequently, we consider if and when one perceives differences in action representation by artificial agents versus humans. Motor resonance and predictive coding models may account for empirical findings, such as an interference effect on action for observed human-like, natural moving characters. However, the expansion of these models to explain more complex behavior, such as empathy, still needs to be investigated in more detail. Finally, we broaden our outlook to social interaction, where virtual reality stimuli can be utilized to imitate complex social situations.

Keywords: uncanny valley, virtual characters, naturalistic stimuli, virtual reality, fMRI, emotion perception, action perception, social interaction

OPEN ACCESS

Edited by:
Marcus Cheetham, University of Zurich, Switzerland

Reviewed by:
Anthony P. Atkinson, Durham University, UK
Lingdan Wu, University of Geneva, Switzerland

****Correspondence:***
Aline W. de Borst, Brain and Emotion Laboratory, Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Oxfordlaan 55, P.O. Box 616, 6200 MD, Maastricht, Netherlands
aline.deborst@maastrichtuniversity.nl

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *25 August 2014*
Accepted: *20 April 2015*
Published: *12 May 2015*

Citation:
de Borst AW and de Gelder B (2015) Is it the real deal? Perception of virtual characters versus humans: an affective cognitive neuroscience perspective. Front. Psychol. 6:576. doi: 10.3389/fpsyg.2015.00576

Introduction

In the last decade, cognitive neuroscience research and especially studies employing brain imaging methods like functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG) underwent significant changes in the type of stimulus material used to investigate human cognition.

Specifically, this was seen in a shift toward using more naturalistic stimuli, as compared to highly controlled, simplified stimuli. For example, Bartels and Zeki (2004) and Hasson et al. (2004), as well as many subsequent investigations, showed new ways to analyze brain activity arising from complex stimulus material such as video clips or entire films (Bartels et al., 2008; Hasson et al., 2008a,b; Nishimoto et al., 2011; Lahnakoski et al., 2012). The developments have given strong impulse and momentum to the fields of social and affective neuroscience in particular, as these fields may profit significantly from the use of naturalistic stimuli. These stimuli are appealing because they have the benefit of being multi-modal, temporally coherent and engaging, and allow for a better understanding of how the brain processes information in complex everyday situations. Importantly, use of such stimuli provides a means to experimentally control the events and interactions to which the participant is exposed, which is difficult to obtain in real social situations (Tikka et al., 2012).

This recent preference for natural stimuli has not been limited to films, but also extended to the domains of robotics and computer-generated (CG) imagery, enabling interdisciplinary groups to branch between these fields of research. For example, MacDorman and Ishiguro (2006) and Chaminade and Cheng (2009) argued for using human-like robots (androids) in cognitive and social science investigations because of their advantages to study human behavior. Similar to films, which can be seen as controllable simulations of reality, MacDorman and Ishiguro (2006) pointed out that androids could be utilized to simulate social situations in a regulated manner. Especially during social interaction, androids have the advantage of physical presence over CG human-like characters. However, CG characters may also be perceived as lifelike, particularly when presented within an immersive three-dimensional virtual environment. Within this virtual environment participants may experience a sense of being there, called "presence" (Slater et al., 2009, 2010). When experiencing high presence, participants respond in a realistic manner to characters and events in the virtual environment (Sanchez-Vives and Slater, 2005). Therefore, CG characters may be a viable alternative to androids in neuroimaging research, since limitations in laboratory set-up (e.g., restrictions due to the magnetic field) rule out the physical presence of androids during fMRI or MEG measurements. Moreover, CG characters are easier to adapt to experimental requirements, the know-how to construct the characters is more widespread, and the costs are lower. However, in those cases in which interaction between the artificial agent and the participant is not needed, both androids and virtual characters may be presented through videos or images as a less technically challenging approach. Altogether, these stimuli are very well suited to study the brain basis of human cognition in a controlled but natural manner. One can come closer to understanding mental processes taking place during planning, social interaction, decision-making, emotion perception and other real-life situations by simulating these activities with virtual stimuli.

However, aside from the many benefits that using virtual characters as stimulus material in neuroscience research may provide, there may also be pitfalls. On many occasions human-like virtual characters are not just seen as representations of humans, but are treated as equivalent to (photo or video material of) humans.

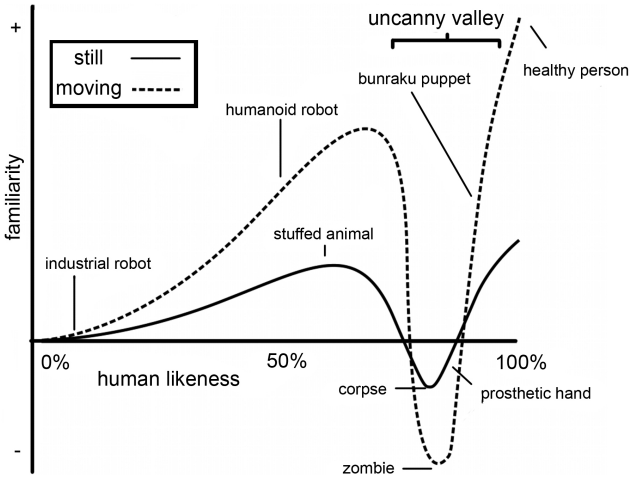


FIGURE 1 | Illustration of the "uncanny valley" function, shown as a simplified version of the figure appearing in Mori's (1970) original Energy article from the translation by MacDorman and Minato (MacDorman, 2005).

For example, in some studies the implications of using CG faces rather than photographs of faces to study human cognitive processes such as emotion perception are not discussed explicitly (a.o. Klasen et al., 2011). This may be problematic, as human-like characters may evoke different behavioral and brain responses than actual humans (as discussed in work from the same group; Sarkheil et al., 2013). This notion was first illustrated with the uncanny valley hypothesis, formulated by robotics professor Mori (1970). He theorized that characters that resemble humans very strongly, but are not human, cause feelings of eeriness in human observers. The uncanny valley hypothesis claims that when the human likeness of a creature increases, the "Shinwakan" (affinity/familiarity) of the creature increases, until a certain point close to 100% human likeness, where a sharp decrease occurs. This decrease is even more pronounced when the creature is moving (see **Figure 1**). This valley in the rise of familiarity, which can be described as an uncanny feeling, is called the "uncanny valley." More recently, this uncanny feeling has been reported during the experience of androids and human-like virtual characters. For example, viewers and actors that were performing alongside the singing human-like robot Geminoid F have described it as eerie and creepy (Waugh, 2012). And the human-like virtual characters in the flopped film production "Final Fantasy: The Spirits Within" were perceived as having "a coldness in the eyes, a mechanical quality in the movements" (Travers, 2001).

However, the differences that may occur between perceiving an artificial agent and a human being are not negative *per se*; they can also teach us more about the conditions under which we still perceive a stimulus as human. In this review we aim to investigate if, when, and how the perception of human-like CG characters and androids differs from the perception of humans, and what factors influence this perception. We approach the matter from a cognitive neuroscience perspective, paying particular attention to the role of virtual characters as potential stimulus material

in social and affective neuroimaging studies, and outlining the differences that the perception of these characters may evoke in underlying brain activity when compared to human stimuli. This investigation partly takes the predictions and properties (human likeness and movement) of the uncanny valley into account, but we extend the discussion to other theories and factors relevant in affective and social neuroscience, such as emotion perception and social interaction. First, we consider how the appearance of a virtual character influences its perception, in which categorization plays an important role. Subsequently, we discuss how emotions expressed by avatars and androids are perceived by human observers. We then briefly touch upon the perception of action, where we discuss if and when one perceives differences in action representation by artificial agents versus humans and consider different theories that may explain why one perceives these differences. Finally, we broaden our outlook and see how these processes interact with other social factors during interaction.

Human Likeness of Virtual Characters

First, we discuss whether the degree of human likeness of virtual characters influences the percept and the feelings they evoke in the observer and whether this gives rise to differences in brain activity. In existing research in this field, human likeness is mainly manipulated through morphing from one image to another in several steps. The human likeness morphing continua generally have one of two endpoints: realistically rendered human-like CG characters (high human likeness) or photographs of humans (100% human likeness). As a starting point a variety of stimuli is used, including non-human CG characters, cartoon characters, robots, or human-like CG characters. The uncanny valley hypothesis predicts that characters with a high human likeness will give rise to a strong sensation of eeriness (**Figure 1**). MacDorman and Ishiguro (2006) claimed to have found empirical evidence for this hypothesis when comparing ratings of human likeness, familiarity, and eeriness for two sets of morphed photographs that ranged from a humanoid robot to an android to a human. They showed a valley in the familiarity rating for the photographs between the humanoid robot and the android, which was accompanied by an increase in eeriness. However, several subsequent studies have suggested that there might be other mechanisms underlying these results. Cheetham and Jancke (2013) pointed out that not one continuum, but several juxtaposed continua were used in MacDorman and Ishiguro's study, leading to discontinuities in the human likeness scale. They suggested that one morph continuum with physically equal steps between morphs is a more unbiased way of investigating the relationship between the dimension of human likeness and other factors. Additionally, Burleigh et al. (2013) discussed two alternative hypotheses that may explain MacDorman and Ishiguro's results: the atypical feature hypothesis and the category conflict hypothesis. The atypical feature hypothesis states that one or more atypical features of the stimulus, such as holes in the forehead that might be typical for a robot but not for a human, together with the level of human likeness account for the perceived eeriness. The category conflict hypothesis suggests that when human likeness of the stimulus is comprised of a morph between two categories (e.g., non-human and human-like) the stimuli in the middle of this scale are perceived as ambiguous, leading to negative affect (Burleigh et al., 2013).

In the first of two studies, Burleigh et al. (2013) manipulated the prototypicality and geometric realism of human-like CG faces in seven equal steps and the participants rated the human likeness, eeriness, fear, disgust, and attractiveness of the faces on 7-point Likert scales. When comparing the subjective human likeness with the perceived eeriness, a linear relation was found between the two, with no presence of an uncanny valley. Similar results were obtained for the relationship between the objective stimulus properties, prototypicality and geometric realism, and the ratings. In a follow-up experiment, the hypotheses mentioned before were tested explicitly, by creating two sets of stimuli. One set manipulated human likeness based on category representation from animal faces to human-looking avatar faces (prototypicality), while the other set varied human likeness on the basis of skin color (blue to natural; realism) for avatars with otherwise human features. An atypical feature (increased eye size) was introduced in both sets. They found that there was a linear relationship between eeriness and subjective human likeness for the realism dataset: when human likeness was high, eeriness was low. The relationship between eeriness and subjective human likeness for the prototypicality dataset was better explained by a quadratic or cubic model, as it showed divergent data points that reflected an increased eeriness around the middle of the subjective human likeness scale. These results supported the category conflict hypothesis. However, the results were not further confirmed by the relationship between objective human likeness and eeriness. Also, no evidence was found for the hypothesis that the combination of the atypical feature with high levels of human likeness increased eeriness.

This categorization conflict has also been investigated in several other studies. Yamada et al. (2013) suggested that categorization difficulty of an ambiguous stimulus leads to higher processing demands and lower processing fluency, which in turn leads to negative affect. They tested morphs between photographs of human and stuffed human, or cartoon human faces (Experiment 1), morphs between photographs of dog and stuffed dog or cartoon dog faces (Experiment 2), and morphs between photographs of different sexes or different identities of human faces (Experiment 3) in forced choice classification and evaluation tasks. The results showed that the most ambiguous images on the realistic versus stuffed/cartoon scales had an increased processing time and showed a decrease in likability, both within and across species. Morphs within one category (sex or identity of human faces) did not show this decrease in likability. The latter results are in line with Burleigh et al. (2013), described above, showing a linear decrease in eeriness with increasing levels of realism (blue to natural skin color) for human-looking avatar faces. This has implications for neuroimaging research in face perception, where within-category morphs are common. These results suggest that no changes in likability should be expected for those continua where human likeness is constant.

In addition to considering stimulus classification speed, Cheetham et al. (2011) also investigated the subjective category boundary for CG avatar to photographic human morphs and measured perceptual discrimination between different sets of these stimuli. They showed that when participants had to classify a

stimulus as either human or avatar in a forced choice task, a sharp category boundary was found, showing a stronger categorization than the physical dissimilarity of the stimuli. For example, when a stimulus was physically 33% human, it was categorized as human 10% of the time, while when a stimulus was 66% human, it was categorized as human 90% of the time. Increased classification response times were found for the most ambiguous stimulus at the estimated category boundary, which is similar to the results obtained by Yamada et al. (2013). In the perceptual discrimination task (see also Cheetham and Jancke, 2013), participants were asked to judge sets of two faces (avatar–human, human–human, avatar–avatar), which had an equal distance on the morphing continuum, as same or different. Participants showed a strong tendency to judge sets of faces that crossed the subjective category boundary as different, while this was significantly less for within category changes. This phenomenon of perceiving within-category differences as smaller than between category differences is typically defined as categorical perception and has been shown for different types of stimulus continua (Harnard, 1987; Liberman, 1996; de Gelder et al., 1997; Looser and Wheatley, 2010). A possible explanation of this phenomenon in general has been described in terms of a Bayesian model by Feldman et al. (2009) and is subsequently incorporated into a model specific to the uncanny valley (for more details, see Moore, 2012). In a subsequent set of studies, Cheetham et al. (2014) tested perceptual discrimination with a different task, which used smaller increments between sets of morphs, and employed another analysis method (d' compared to A' and % different responses). The task (ABX task; Liberman et al., 1957) required participants to judge to which of two previously shown faces a target face corresponded. In the first study, which tested the pattern of perceptual discrimination performance along the dimension of human likeness, discrimination sensitivity of within-category avatar faces was increased compared to across-category faces and within-category human faces. The within-category human faces showed the lowest discrimination sensitivity. In the second experiment, the morphing direction was switched compared to Experiment 1 in order to eliminate possible influences of the morphing algorithm on the perceptual discrimination pattern. In Experiment 2, within-category avatar faces and across-category faces showed increased levels of perceptual discrimination compared to within-category human faces. Participants also rated familiarity of the images, which—in line with other studies—increased with increasing human likeness. Individual variability in perceptual discrimination correlated negatively with familiarity ratings for avatar and ambiguous faces. Finally, Cheetham et al. (2014) investigated in Experiment 3 whether their results could be explained by a differential processing bias, e.g., within-group humans are processed at the exemplar level, while out-group avatars are processed at the category level. This seemed not to be the case, as inversion of the stimuli produced the same results as Experiment 2. Overall, the results indicate that perceptual discrimination was asymmetrical along the human likeness dimension (from human-like to human), with lower discrimination sensitivity for human faces, and that familiarity increased with human likeness. Cheetham et al. (2014) related these findings to fluency amplification (Albrecht and Carbon, 2014): higher levels of fluency (enhanced discrimination) went together with amplified affect (higher feelings of strangeness). However, as processing fluency is broadly defined as "the ease with which information is processed," various tasks may measure and highlight different aspects of ease of information processing.

Accompanying brain data (Cheetham et al., 2011) suggested that the brain encodes physical and categorical changes of the stimulus differently. The results showed that mid-fusiform regions responded to physical change, while the medial temporal lobe (MTL) and a number of subcortical regions were sensitive for category change. The fusiform face area (FFA) is not only seen as a region particularly responsive to faces, but as responsive to any fine grained distinctions between expertise-acquired categories (Gauthier et al., 2000). Therefore, the activation of fusiform gyrus to the physical change of the morphed stimuli may be understood within this context, as discrimination of face-features depends on experience. Moreover, the physical change in human faces activated a more extensive network of brain regions than physical change in avatar faces. The category change from avatar to human mainly activated the MTL, amygdala and insula, while subcortical regions responded to change from human to avatar. The authors related the MTL activation for human targets to category processing and learning, suggesting that different categorization problems underlie avatar and human target faces. Generally, the subcortical regions and insula that were activated for category change may be related to processing the novelty and uncertainty of across-category stimuli. However, since the insula has been shown to be involved in a wide variety of tasks (ranging from disgust to mental imagery to conflict monitoring), more specific hypotheses should be tested in order to better understand the role of the insula. Moreover, subsequent neuroimaging research could also further validate the other fMRI findings of this study, especially targeting the function of the subcortical regions, such as the thalamus and putamen, in avatar target perception.

In another study Cheetham et al. (2013) examined eye-tracking data during the forced choice categorization task described above. The results confirmed that the most ambiguous stimulus at the subjective category boundary generated the largest conflict in decision making as reflected by increased response times. For ambiguous faces compared to unambiguous avatar faces the dwell time (i.e., duration of the fixation) on the eyes and mouth increased. Thus, the relative importance of facial features changed depending on the category ambiguity. Compared to human faces this did not reach significance.

McDonnell et al. (2012) took both the human-likeness and motion parameters of the uncanny valley hypothesis into account in their study. They created a face-continuum from an abstract cartoon character to a highly realistic human-like CG character. These virtual faces were studied as still images as well as moving images, which were animated using human motion capture. The characters were judged on different aspects, such as realism, familiarity, and appeal. The drop in appeal was not found for the most realistic stimuli, as the uncanny valley hypothesis would predict, but rather for those on the border between cartoon-like and realistic. Consistent with some of the previously discussed studies, McDonnell et al. (2012) suggested that this might be caused by the fact that the characters in the middle of the abstract-to-realism scale may be more difficult to categorize, especially

with the mismatch between appearance and motion. Movement of the characters amplified the effects of appeal (higher for appealing stimuli, lower for non-appealing stimuli), which is in line with the uncanny valley hypothesis (see **Figure 1**). Movement is only one of the properties that can enhance feelings of eeriness when it's misaligned with the visual appearance of the stimulus. Mitchell et al. (2011) created a cross-modal mismatch between human likeness of a face and the corresponding voice. The incongruent conditions (human face with synthetic voice or robot face with human voice) showed increased ratings of eeriness compared to their congruent counterparts.

Thus, subjective categorization seems to influence processing and experience of stimuli along the dimension of human likeness, when these stimuli are morphed across different categories. This held true for morphing continua from non-human to human-like as well as human-like to human. Ambiguous stimuli at the subjective category boundaries gave rise to prolonged classification response times. However, in other tasks such as perceptual discrimination, processing of ambiguous stimuli may instead be facilitated and differences between avatar and human stimulus processing were observed. In some cases, especially when the morphing continua ranged from non-human to human-like, the most ambiguous stimuli increased eeriness ratings. For human-like to human continua eeriness seemed to decrease linearly. Morphing of faces between identities or genders with equal human likeness showed little changes in likability. However, expertise in human face discrimination among humans may influence processing of human faces versus avatar faces and modulate underlying brain regions (e.g., the fusiform gyrus) that respond to the physical properties of the stimulus. This may have consequences for neuroimaging studies that use avatar faces to study face perception even when comparing avatar conditions directly, as the underlying mechanisms that give rise to enhance or reduced activity to avatar faces are not fully understood. Given the limited number of studies on this subject, an expansion of neuroimaging studies that compare different properties of avatar and human faces and investigate its underlying brain activity in an experimentally and statistically well-constructed way is needed in order to further understand its underlying neuronal mechanisms. Moreover, when designing multi-modal stimuli, the movement and auditory components should match the human likeness (e.g., high human likeness with natural movement and human voice) in order to avoid the sensation of eeriness.

Perception of Emotion in Virtual Characters and Robots

In the previous section we discussed the feelings that arise from perceiving neutral avatar or human faces with different levels of human-likeness. However, in the field of affective neuroscience the emphasis is not so much on whether neutral stimuli evoke emotions, but rather on how emotional stimuli are perceived. Therefore, we extend the comparison between avatar and human stimuli to emotional faces and bodies.

Several behavioral studies compared the perception of affective human-like avatar faces with the perception of human faces using either still or moving images. For example, Rizzo et al. (2012) compared video clips of facial expressions of emotions by humans or 3D CG avatars. Different types of emotions were expressed, including the six universal emotions (happy, sad, fear, anger, disgust, surprise). Participants were asked to indicate which type of emotion was expressed and how much the clip expressed each of the emotions. The results indicated that the emotions were equally convincing for avatar and human faces. However, the percentage of the clips that were correctly categorized differed for avatar and human video conditions. These results were not very consistent, e.g., sometimes avatars were more correctly identified, while at other times human clips were more correctly categorized. This seemed to depend strongly on the actor used to express the emotion. Another study showed that photographs of human faces and human-like CG avatar faces were recognized comparably well, but that recognition differences occurred for specific emotions (Dyck et al., 2008). For example, disgust was recognized less well, while sadness and fear were recognized better in avatar faces compared to human faces. Thus, behaviorally, recognition of emotions seems to rely more strongly on how (well) the emotions are expressed rather than whether they are expressed by an avatar or human face. However, beyond the explicit recognition of emotions, it is interesting to understand whether emotions expressed by avatars or humans evoke similar patterns in motor and brain responses. First, we review research on facial expressions of emotion within this context and subsequently discuss bodily emotions.

Motor Responses to Affective Virtual Faces

When observing emotional faces, small responses in the facial muscles of the viewer take place in those muscles that are used for the expression of the emotion. For example, viewing happy human faces is accompanied with electromyography (EMG) activity in the zygomaticus major (ZM), the main muscle for expressing a smile, and viewing angry human faces evokes activity in corrugator supercilii (CS), the muscle for expressing a frown (Dimberg and Petterson, 2000; Dimberg et al., 2000; Aguado et al., 2013). It has been shown that perceiving emotional CG avatar faces results in EMG activity in the same facial muscles as perceiving photographs of human faces, e.g., the ZM for happy avatar faces (Weyers et al., 2006; Likowski et al., 2012) and the CS for sad and angry avatar faces (Likowski et al., 2012). This implies that viewing emotional avatar faces evokes the same muscle responses—called mimicry—in humans as viewing emotional human faces. Dynamic avatar faces (morphed from neutral to an emotion) showed increased EMG activity for happy faces, compared to neutral faces, but this did not extend to angry faces in a study by Weyers et al. (2006). For their avatars the CS activity for dynamic and static angry faces was not significantly different from CS activity for neutral faces. In another study Weyers et al. (2009) showed that these mimicry effects are susceptible to subconscious priming, suggesting that subconscious motives influence empathic mimicry. After priming with neutral words (e.g., street) facial mimicry of emotional avatar faces occurred, but this effect was reduced (less relaxation of CS for happy faces) when the participants were primed with competitive words such as rival or opponent. Likowski et al. (2012) showed that the congruent facial responses between observer and avatar correlated with brain

activity in a large network of regions, including inferior frontal gyrus and inferior parietal lobe, regions that have been shown as part of the mirror neuron system. Mirror neurons are neurons that are active during motor execution, but also respond to action observation (Gallese et al., 1996). These results further supported the notion that emotional avatars evoke mimicking behavior in humans and that this is accompanied by activation of brain regions that also activate for the expression of these emotions. In the next paragraph and the section on action perception we will further discuss their role and relevance to action and emotion perception in avatars.

Brain Responses to Affective Virtual Faces and Robots

We can further investigate the effects of human likeness on emotion perception by looking at accompanying brain activity. Moser et al. (2007) compared brain activity for the recognition of emotions in CG avatar faces versus photographs of human faces. Behaviorally, female participants showed better recognition of emotions in human faces compared to avatar faces in a forced choice task, while males showed no differences. When looking at the brain data for the group as a whole, human and avatar faces evoked similar activity in the amygdala. These results suggest that animated faces may be as effective to investigate perception of emotional facial expressions as human faces. However, these findings are not entirely consistent with the behavioral results, where—at least for women—differences were found in emotion recognition between the two types of stimuli. Therefore, it would be interesting to repeat the experiment with a larger group of subjects, to see if the behavioral differences between the sexes also translate in differential amygdala activation. When comparing the two conditions (human faces versus avatar faces) directly, differences were found in the fusiform gyrus. Previous research has shown that perception of human faces activates the FFA more strongly than other faces, such as animal faces (Kanwisher et al., 1999). The differences in activation found by Moser et al. (2007) might be caused by the fact that the FFA is driven more strongly by within-species faces, as humans are most experienced with classifying human faces. These initial results seem to suggest that although physical differences are perceived between avatar and human faces, the expressed emotions may still be processed in a similar manner.

Opposite effects on the modulation of brain activity in the fusiform gyrus were found for the comparison of emotional facial expressions by a mechanical robot versus a human (Chaminade et al., 2010). Viewing videos of the robot evoked stronger responses overall in visual areas V3, V4, V5, and FFA. Perhaps, because the features of the robot face were so different from the human or avatar face, more visual processing was required to recognize the robot face, leading to enhanced activity in these regions. When looking at the different emotions, emotion-specific activations were found in the insular cortex for disgust and in the right putamen for joy. Although the activations were reduced for the robot emotional expressions, they were not significantly different from the brain responses to the human faces for these emotions. The participants did show enhanced brain activity in orbitofrontal cortex for angry human stimuli compared to angry robot stimuli (which did not differ from baseline). The reduced response to the angry robot stimuli may stem from the fact that the avatar angry faces were rated as significantly less angry than the human angry faces. Chaminade et al. (2010) interpret their results for the insula and putamen in the context of motor resonance, a reaction that has been suggested to rely on mirror neurons. They propose that for viewing human and avatar emotional faces resonance occurs in the observer, which may then play a role in understanding the other person. In the decades since its introduction, the notion of a mirror neuron system at the basis of motor perception has been expanded to explain complex behaviors such as imitation, emotion observation, intention, and empathy (Rizzolatti et al., 2001; Wicker et al., 2003; Iacoboni et al., 2005). While strong evidence has been found for mirror neurons in the context of motor observation, as discussed in the next section, the roles of the regions activated for the experience and observation of emotions and empathy are less clear. For emotions with basic underlying mechanisms such as pain, mirroring properties might hold true, but the more complex the emotional process, the more other mechanisms might come into play. For example, neuroimaging research since the late 19s has shown that when we imagine objects, places or voices, similar regions activate as when we perceive these categories (Cohen et al., 1996; Mellet et al., 1996; Ishai et al., 2000; O'Craven and Kanwisher, 2000; Trojano et al., 2000; Formisano et al., 2002; de Borst et al., 2012). In a sense, imagery also makes the regions involved in perception resonate (Kilner et al., 2007b). As imagery might play a significant role in processes such as empathy, and there is partial overlap between regions attributed to the mirror neuron system and mental imagery networks, it is difficult to attribute the brain activity to mirror neurons *per se*.

In conclusion, for faces looking very dissimilar from human faces, the expressed emotions may evoke reduced responses in the observers, as expressed by lower intensity ratings and reduced brain activity. When emotional avatar faces look highly similar to human faces, they may evoke similar emotional responses as expressed by mimicking responses in the face and activation of emotion regulatory regions. However, differences in brain activity still may occur as a response to the physical differences between avatar and human stimuli. This may be caused by the experience people have with viewing and interpreting human faces.

Multi-Sensory Integration of Virtual Bodies and Voices

When presenting multi-modal affective virtual stimuli, not only the interaction between human likeness of appearance, movement and voice comes into play, but also the congruency of the expressed emotional content. For multi-modal affective human stimuli, de Gelder et al. (1999) have first shown that affective facial expressions and emotional voice prosody influence each other. A follow-up study by de Gelder and Vroomen (2000a) showed that emotional categorization of facial emotional expressions along a morph continuum from sad to happy was biased toward the emotion expressed in the simultaneously presented voice and *vice versa*. Similarly, Stienen et al. (2011) have shown that emotional human body postures and emotional human voices influenced each other, even when the participants were unaware of the

bodies. More recent work has shown that the conscious categorization of ambiguous, affective videos of human bodies (Watson and de Gelder, 2014) and CG human-like bodies (de Gelder et al., 2014) was also influenced by the emotion of human voices. Classification of emotions expressed by CG avatar bodies that were morphed on a continuum from happy to angry, showed an inverted u-shape for response times when participants judged emotion visually. The emotionally ambiguous stimuli showed the largest response times. The categorization curve for emotional avatars showed an increasing percentage of anger responses with the gradual shift from happy to angry. For bi-modal stimuli consisting of a simultaneous CG body and voice expression, voices influenced the rating of the bodily expression in the morphed continuum (de Gelder et al., 2014), consistent with studies on human multi-sensory integration between emotional faces and voices (de Gelder and Vroomen, 2000b; Campanella and Belin, 2007). These initial results on multi-sensory integration of affective virtual bodies and voices indicate that the behavioral effects are similar to those observed with human bodies and voices.

Emotion Perception in Virtual Reality

Outside of the laboratory, emotion perception often occurs in more complex situations. Some studies have tried to investigate these situations in social experiments. One such example is the famous Milgram experiment on obedience to authority figures, in which participants were instructed to give what they believe are painful electric shocks to another participant each time that participant answers wrongly during a task (Milgram, 1963). Even though performing the shocks gave great distress to the participants, more than half of the participants continued to do so until a final 450-volt shock. Slater et al. (2006) showed that during a 3D virtual version of this experiment participants reacted behaviorally and physiologically as if it were real, even though they knew it was not. Functional MRI evidence showed that individual differences in personal distress during the virtual Milgram experiment co-vary with neuronal changes during perception of the avatar in pain, while no covariance was found with individual changes in emphatic concern (Cheetham et al., 2009). Other life-like responses to virtual reality situations were shown in a bystander study (Slater et al., 2013), in which football supporters were more likely to physically intervene in a confrontation when their attacked CG conversation partner was from the same football club. These results replicate earlier findings from choreographed human situations and illustrate how virtual stimuli can be utilized to imitate complex social situations that might be difficult to orchestrate otherwise.

Evidence so far seems to suggest that expressions of emotions in virtual characters can be perceived similarly to human emotion, with corresponding behavioral and physiological activation. In the brain, evidence for this further accumulates, as emotion-specific regions show similar activation for human-like artificial agents and humans, although physical dissimilarities are also visible. Some typical brain mechanisms, such as multi-sensory integration, seem to influence emotional avatar perception in a manner comparable to the perception of emotions in humans. However, multi-sensory integration is not a phenomenon that occurs only with the perception of humans, but rather is a more general mechanism for integration of sensory modalities in the brain.

Action Perception in Virtual Characters and Robots

As already mentioned previously, the movement of virtual characters also influences the way in which they are perceived. The interaction between movement and appearance of a human-like stimulus, in behavioral and neuronal effects, has been interpreted in the context of several relatable theories, stemming from different fields. We will briefly discuss these theories and review their empirical support in the current context. The uncanny valley hypothesis, that focusses on behavioral effects, suggests that adding movement increases the familiarity for stimuli that were rated as likeable when still, e.g., for characters with extremes of human likeness on the left and right side of the uncanny valley (see **Figure 1**). Movement decreases the familiarity even further for human-like images that were rated as unlikeable when still. Comparisons between human movement and avatar/robot movement have also been made in the context of motor resonance. In humans (as well as in the monkey), mirror neurons activate both to the execution and observation of motor actions. Since these neurons are seen as a way to predict and infer actions, robots and virtual characters are quite suitable to study whether our brain only resonates for observing human actions, since these resemble our own motor system, or whether this also occurs for mechanical and CG actions. Some supporters of this theory would predict more resonance (e.g., activity in motor regions) for human-like than artificial action stimuli. In this review, we only briefly touch upon this subject. For more elaborate reviews on the role of mirror neurons and resonance in the perception of androids see Chaminade and Cheng (2009) and Sciutti et al. (2012). Finally, the predictive coding model (Friston, 2005, 2010; Kilner et al., 2007a) suggests that the brain tries to optimize processing at all levels of the cortex, by integrating bottom-up and top-down information through recurrent, reciprocal interactions. At each level predictions are made of the representation in the level below. Through these interactions the error between the sensory expression and its cause is minimized. This framework has been used as a way to explain motor resonance (Kilner et al., 2007a,b). Also, some authors have used this model as an explanation for the uncanny valley interaction of movement and appearance of human-like characters (Saygin et al., 2011). Although predictive coding is well-described for, e.g., action perception and observation, where links between cause (motor goals) and sensory expression (observed kinematics) are relatively direct, generalizing this model to more intricate social phenomena might be more complicated.

Interaction of Motion and Appearance in Virtual Characters

While the influence of movement on the perception of virtual characters within the uncanny valley hypothesis was confirmed by McDonnell et al. (2012), no such effect was found by Thompson et al. (2011). In their study they manipulated the gait of a human-like CG character and an abstract mannequin CG character based

on three kinematic features: articulation, phase, and jerk. The results showed that ratings of humanness and familiarity increased monotonically from least natural to most natural for each of the three kinematic features. An opposite pattern, that is decrease, was found for ratings of eeriness. No differences were found between the mannequin and the human-like avatar. Thus, changes in movement parameters did not show an uncanny valley effect. However, the human likeness parameter was not parametrically adapted in this study. Therefore, it makes these results difficult to compare to those of McDonnell et al. (2012). Ideally, both the human likeness and the kinematic parameters should be manipulated and compared to get a full understanding of the phenomenon. Piwek et al. (2014) manipulated these two parameters and found evidence for the uncanny valley in human likeness, but added motion only increased acceptability of the stimulus, no matter if it was natural or distorted. Their results are in line with Thompson et al. (2011), showing improved familiarity or acceptability when avatars were moving instead of still.

The interaction between appearance and motion can also be investigated in the opposite direction: not the influence of motion on the rating of the appearance, but the influence of appearance on the rating of naturalness of the motion. Chaminade et al. (2007) showed that the response bias to rate a character as biological depends on its human likeness, where higher human likeness coincides with lower ratings of "biological" for both human motion capture data and animated data. However, in this study the degree of human likeness of the virtual characters was not equally spaced. For example, the monster gave a similar response bias to "biological" as the human-like jogger.

Motion Perception of Artificial Agents in the Brain

Research on the neuronal basis of motion perception has shown differences for avatar or robot motion perception versus human motion perception, as well as congruency effects for combinations of artificial and biological motion with human likeness of the character. For example, the perception of human grasping actions activates the premotor cortex, while the same actions performed by a robot arm do not (Tai et al., 2004). This is in line with motor resonance theory and with results from other behavioral studies (Kilner et al., 2003; Press et al., 2005). These studies showed that executing an arm movement is interfered by observing a human performing an incongruent movement, while this congruency effect does not occur (Kilner et al., 2003), or to a smaller extent (Press et al., 2005) when observing a robot performing an incongruent movement. This effect has been suggested to originate from the velocity profile of biological motion (Kilner et al., 2007c) and interacts with previous experience (Press et al., 2007). However, when the robot has both a human-like appearance and moves naturally, this congruency effect for movement can be found for both robot and human movements (Oztop et al., 2005). These results suggest that when a robot is human-like, motor resonance occurs. In line with previously discussed behavioral findings (Chaminade et al., 2007; McDonnell et al., 2012), an fMRI study by Saygin et al. (2011) showed that when appearance and the expected nature of motion do not match, distinct responses appear in the brain. When investigating repetition suppression in the brain (the reduction of neuronal responses for repeated presentation of the stimulus) for passive viewing of videos with natural human biological motion, videos of robots with artificial motion or videos of humanoids with artificial motion, the largest and most wide-spread suppression effects were found for the incongruent stimulus (i.e., the android with artificial motion), especially in the anterior intraparietal sulcus. This effect seemed to be caused by stronger initial activity (unrepeated stimulus) for the android compared to the robot and human. Saygin et al. (2011) interpret their results on the basis of the predictive coding model, as an increased prediction error in the brain when having to conciliate a human-like character with non-biological motion properties. However, they do not specify how this interaction between properties of different senses (human likeness in the visual domain and naturalness of motion in the motor domain) would be explained by the predictive coding model. The integration across senses and the generation of affective states in the context of the predictive coding model has been discussed more recently in studies on emotion perception, self-representation and multi-sensory integration (Seth, 2013; Apps and Tsakiris, 2014; Ishida et al., 2014; Sel, 2014).

It is fair to conclude that perceived human likeness of a virtual character or robot varies with the naturalness of motion, where high human likeness combined with artificial motion shows an incongruency effect which might be caused by a prediction error in the brain that can be related to higher levels of eeriness experience. The prediction error occurs when two properties of the stimulus do not match and for action perception the prediction error could occur in the mirror neuron system. This suggests that it is important to animate human-like virtual characters and program robots with human-like motion data. Several EMG studies by Huis in 't Veld et al. (2014a,b) suggested that specific muscle groups are used for the bodily expressions of emotion in humans. This information, together with motion capture data could be used to improve modeling of biological movements for virtual agents and robots.

Human–Avatar Social Interaction

The human likeness, naturalness of movement and emotions expressed and evoked by a virtual character or robot are important factors influencing their perception. These and other social factors become particularly relevant when avatars and androids interact with humans. Therefore, in this last section we will go beyond what has been discussed so far and consider implicit processing of virtual characters and robots in order to understand more about social interaction between humans and artificial agents.

In an experiment by McDonnell et al. (2012), participants were asked to tell if a virtual character was lying or telling the truth. There were three CG characters, each rendered differently (cartoon, semi-realistic, highly-realistic) combined with an audio track. As a control, the audio was presented by itself or together with a video of the motion capture session (human). The authors expected that the more unappealing characters would bias the participants toward thinking that they lie more. No such effects were found. The lie ratings and the bias toward believing them were similar for the different renderings and the videos of real humans. However, participants may have extracted most of their

information from the audio track that was identical across all stimuli.

When interacting directly with a virtual character, people often show behavior that is similar to human interaction. For example, when offering a gift to a virtual character, participants react much in the same way as they would with humans when this gift is either accepted or rejected (Zucker et al., 2011). When being rejected, brain activity in the anterior insula increases. As discussed previously, the insular cortex has been shown to be involved in a wide variety of tasks, most relating to subjective feelings (Craig, 2009). When the facial expression (happy/disgust) of the virtual character is incongruent with the hand movement (accepting/rejecting), activity in the superior temporal sulcus (STS) rises. This is in line with another study (Vander Wyck et al., 2009) that showed STS activation for viewing of incongruent actions by a human actress (picking up object) based on the emotional context (negative regard), suggesting its role in perception of social acts. Moreover, work by Slater, Sanchez-Vives, and others on virtual embodiment showed that interacting with virtual characters in virtual reality while being embodied in a virtual character gives rise to specific character-dependent changes in behavior, ranging from pain perception to implicit racial bias (a.o. Banakou et al., 2013; Llobera et al., 2013; Peck et al., 2013; Martini et al., 2014). These results indicate that virtual reality stimuli can be utilized to imitate complex social situations and may also affect behavior.

Lucas et al. (2014) even take human-virtual character interaction to another level by suggesting that in some particular cases of social interactions virtual characters might be more successful than real humans. They showed that when disclosing health information, participants were more willing to disclose information to an automated virtual character than to a virtual character controlled by a human operator. They had less hesitancy to disclose information and showed their emotions more openly to the automated virtual character.

Obviously, these interactions are not the only factors influencing how humans perceive artificial agents. Robots and avatars can be programmed to display distinct personalities, and these personalities influence whether they are liked or not. For example, Elkins and Derrick (2013) showed that a virtual interviewer is trusted more when the pitch of the voice is lower during the start of the interview and if the avatar smiles. However, how a virtual characters' personality is perceived depends largely on context and the personality of the perceiver. When a robot takes on the role of a healthcare assistant, it should have a different personality than when it works as a security guard. People preferred to have an extraverted healthcare robot, showing greater affect, more positive attitudes and greater trust, compared to an introverted robot (Tay et al., 2014). For the security guard however, people showed the opposite preference—it was perceived better, as more trustworthy and more in control when having the introverted personality. The preference for one or the other humanoid personality does not depend only on situation or the task, but also is a function of the participants' own personality. Extraverts seemed to prefer an extraverted humanoid to encourage them during rehabilitation, while the introverts favored a more nurturing personality (Tapus et al., 2008). Robot or avatar personalities thus may be taken into account when designing stimuli for social neuroimaging experiments. In combination with personality questionnaires for the participant and other hypothesis-relevant measures, personality profiling of avatars may be especially advantageous for virtual reality experiments.

Conclusion

When designing human-like characters to investigate human cognition in neuroimaging research evidence so far indicates that, contradictory to the predictions of the uncanny valley hypothesis, the most human-like characters are processed most similarly to human stimuli on a behavioral and neuronal level. Thus, not the most realistic looking virtual characters evoke an eerie feeling, but rather those on the border between non-human and human categories, especially if they are combined with human-like motion. This subjective experience seems to arise from difficulty in categorizing ambiguous characters that look neither human nor robot-, avatar- or animal-like, which also leads to increased response times for categorization.

Since humans are experienced in perceiving human faces, viewing avatars may evoke differential processing, e.g., enhanced perceptual discrimination, and modulate underlying brain activity. This suggests that results from avatar and human data may not always be comparable and should be interpreted with care. The underlying mechanisms that give rise to this modulation are not fully understood and therefore further neuroimaging research that compares different physical properties of avatars and humans is needed.

The perception of emotional expressions by human-like characters seems to be fairly similar to perception of human emotions, with corresponding behavioral and physiological activation. This is supported by brain data, although again the physical properties of the stimuli may still cause neuronal differences. For artificial faces that look very dissimilar from human faces, the expressed emotions may evoke reduced responses in the observers, as expressed by lower intensity ratings and reduced brain activity.

Research on the influence of movement on the perceived eeriness of artificial characters shows conflicting results. One study showed the described uncanny valley effect, with an additive effect of motion, while other studies found that added motion only increased the familiarity of the characters. Human-like avatars that move realistically are more likeable and perceived as similar to real humans, as shown, e.g., by the behavioral motion-interference effect and motor resonance in the brain. Non-realistic avatars or robots do not show these effects. Eerie feelings for human-like characters with artificial motion might be explained by the predictive coding model, when the predicted human motion patterns and observed artificial motions lead to an increased prediction error. However, this still needs to be investigated in more detail in order to elaborate the model to more complex processes. It is important in social neuroscience research that, when moving avatars or robots are used, their motion is modeled with biologically appropriate parameters and that possible perceptual differences are taken into account.

When socially interacting with humanoids people may perceive and react as if they were interacting with human beings, showing brain activity in regions relating to emotion and interpersonal

experience. Virtual reality experiments may play a significant role in simulating social situations, as these have shown to directly affect social behavior. Neuroimaging experiments could further investigate these virtual experiences by measuring the specific neuronal modulations that lay at the foundation of the behavioral responses.

Acknowledgments

The authors thank Prof. Tapio Takala, Prof. Niklas Ravaja, Dr. Jari Kätsyri, Dr. Pia Tikka, Klaus Förger, and Meeri Mäkäräinen for fruitful discussions on this topic. AdB and BdG were supported by FP7/2007-2013, ERC grant agreement number 295673.

References

Aguado, L., Román, F. J., Rodríguez, S., Diéguez-Risco, T., Romero-Ferreiro, V., and Fernández-Cahill, M. (2013). Learning of facial responses to faces associated with positive or negative emotional expressions. *Span. J. Psychol.* 16, E24. doi: 10.1017/sjp.2013.31

Albrecht, S., and Carbon, C. C. (2014). The Fluency Amplification Model: fluent stimuli show more intense but not evidently more positive evaluations. *Acta Psychol.* 148, 195–203. doi: 10.1016/j.actpsy.2014.02.002

Apps, M. A. J., and Tsakiris, M. (2014). The free-energy self: a predictive coding account of self-recognition. *Neurosci. Biobehav. Rev.* 41, 85–97. doi: 10.1016/j.neubiorev.2013.01.029

Banakou, D., Groten, R., and Slater, M. (2013). Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12846–12851. doi: 10.1073/pnas.1306779110

Bartels, A., and Zeki, S. (2004). Functional brain mapping during free viewing of natural scenes. *Hum. Brain Mapp.* 21, 75–85. doi: 10.1002/hbm.10153

Bartels, A., Zeki, S., and Logothetis, N. K. (2008). Natural vision reveals regional specialization to local motion and to contrast-invariant, global flow in the human brain. *Cereb. Cortex* 18, 705–717. doi: 10.1093/cercor/bhm107

Burleigh, T. J., Schoenherr, J. R., and 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. *Comput. Hum. Behav.* 29, 759–771. doi: 10.1016/j.chb.2012.11.021

Campanella, S., and Belin, P. (2007). Integrating face and voice in person perception. *Trends Cogn. Sci.* 11, 535–543. doi: 10.1016/j.tics.2007.10.001

Chaminade, T., and Cheng, G. (2009). Social cognitive neuroscience and humanoid robotics. *J. Physiol.* 103, 286–295. doi: 10.1016/j.jphysparis.2009.08.011

Chaminade, T., Hodgins, J., and Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters' actions. *Soc. Cogn. Affect. Neurosci.* 2, 206–216. doi: 10.1093/scan/nsm017

Chaminade, T., Zecca, M., Blakemore, S. J., Takanishi, A., Frith, C. D., Micera, S., et al. (2010). Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures. *PLoS ONE* 5:e11577. doi: 10.1371/journal.pone.0011577

Cheetham, M., and Jancke, L. (2013). Perceptual and category processing of the uncanny valley hypothesis' dimension of human likeness: some methodological issues. *J. Vis. Exp.* 76, e4375. doi: 10.3791/4375

Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., and Jancke, L. (2013). Category processing and the human likeness dimension of the uncanny valley hypothesis: eye-tracking data. *Front. Psychol.* 4:108. doi: 10.3389/fpsyg.2013.00108

Cheetham, M., Pedroni, A. F., Antley, A., Slater, M., and Jäncke, L. (2009). Virtual milgram: empathic concern or personal distress? Evidence from functional MRI and dispositional measures. *Front. Hum. Neurosci.* 3:29. doi: 10.3389/neuro.09.029.2009

Cheetham, M., Suter, P., and Jäncke, L. (2011). The human likeness dimension of the "uncanny valley hypothesis": behavioural and functional MRI findings. *Front. Hum. Neurosci.* 5:126. doi: 10.3389/fnhum.2011.00126

Cheetham, M., Suter, P., and Jäncke, L. (2014). Perceptual discrimination difficulty and familiarity in the uncanny valley: more like a " Happy Valley". *Front. Psychol.* 5:1219. doi: 10.3389/fpsyg.2014.01219

Cohen, M. S., Kosslyn, S. M., Breiter, H. C., DiGirolamo, G. J., Thompson, W. L., Anderson, A. K., et al. (1996). Changes in cortical activity during mental rotation. A mapping study using functional MRI. *Brain* 119, 89–100. doi: 10.1093/brain/119.1.89

Craig, A. D. (2009). How do you feel—now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70. doi: 10.1038/nrn2555

de Borst, A. W., Sack, A. T., Jansma, B. M., Esposito, F., de Martino, F., Valente, G., et al. (2012). Integration of "what" and "where" in frontal cortex during visual imagery of scenes. *Neuroimage* 60, 47–58. doi: 10.1016/j.neuroimage.2011.12.005

de Gelder, B., Böcker, K. B., Tuomainen, J., Hensen, M., and Vroomen, J. (1999). The combined perception of emotion from voice and face: early interaction revealed by human electric brain responses. *Neurosci. Lett.* 260, 133–136. doi: 10.1016/S0304-3940(98)00963-X

de Gelder, B., de Borst, A. W., and Watson, R. (2014). The perception of emotion in body expressions. *WIREs Cogn. Sci.* 6, 149–158. doi: 10.1002/wcs.1335

de Gelder, B., Teunisse, J.-P., and Benson, P. J. (1997). Categorical perception of facial expressions: categories and their internal structure. *Cogn. Emot.* 11, 1–23. doi: 10.1080/026999397380005

de Gelder, B., and Vroomen, J. (2000a). The perception of emotions by ear and by eye. *Cogn. Emot.* 14, 289–311. doi: 10.1080/026999300378824

de Gelder, B., and Vroomen, J. (2000b). Bimodal emotion perception: integration across separate modalities, cross-modal perceptual grouping or perception of multimodal events? *Cogn. Emot.* 14, 321–324. doi: 10.1080/026999300378842

Dimberg, U., and Petterson, M. (2000). Facial reactions to happy and angry facial expressions: evidence for right hemisphere dominance. *Psychophysiology* 37, 693–696. doi: 10.1111/1469-8986.3750693

Dimberg, U., Thunberg, M., and Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychol. Sci.* 11, 86–89. doi: 10.1111/1467-9280.00221

Dyck, M., Winbeck, M., Leiberg, S., Chen, Y., Gur, R. C., and Mathiak, K. (2008). Recognition profile of emotions in natural and virtual faces. *PLoS ONE* 3:e3628. doi: 10.1371/journal.pone.0003628

Elkins, A. C., and Derrick, D. C. (2013). The sound of trust: voice as a measurement of trust during interactions with embodied conversational agents. *Group Decis. Negot.* 22, 897–913. doi: 10.1007/s10726-012-9339-x

Feldman, N. H., Griffiths, T. L., and Morgan, J. L. (2009). The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychol. Rev.* 116, 752–782. doi: 10.1037/a0017196

Formisano, E., Linden, D. E., Di Salle, F., Trojano, L., Esposito, F., Sack, A. T., et al. (2002). Tracking the mind's image in the brain I: time-resolved fMRI during visuospatial mental imagery. *Neuron* 35, 185–194. doi: 10.1016/S0896-6273(02)00747-X

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain* 119, 593–609. doi: 10.1093/brain/119.2.593

Gauthier, I., Skudlarski, P., Gore, J. C., and Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nat. Neurosci.* 3, 191–197. doi: 10.1038/72140

Harnard, S. (1987). "Psychophysical and cognitive aspects of categorical perception: a critical overview," in *Categorical Perception: The Groundwork of Cognition*, ed. S. Harnad (New York, NY: Cambridge University Press), 1–25.

Hasson, U., Furman, O., Clark, D., Dudai, Y., and Davachi, L. (2008a). Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. *Neuron* 57, 452–462. doi: 10.1016/j.neuron.2007.12.009

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., and Rubin, N. (2008b). A hierarchy of temporal receptive windows in human cortex. *J. Neurosci.* 28, 2539–2550. doi: 10.1523/JNEUROSCI.5487-07.2008

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science* 303, 1634–1640. doi: 10.1126/science.1089506

Huis in 't Veld, E. M., Van Boxtel, G. J., and de Gelder, B. (2014a). The Body Action Coding System I: muscle activations during the perception and expression of emotion. *Soc. Neurosci.* 9, 249–264. doi: 10.1080/17470919.2014.890668

Huis in 't Veld, E. M., Van Boxtel, G. J., and de Gelder, B. (2014b). The Body Action Coding System II: muscle activations during the perception and expression of emotion. *Front. Behav. Neurosci.* 8:330. doi: 10.3389/fnbeh.2014.00330

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., and Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol.* 3:e79. doi: 10.1371/journal.pbio.0030079

Ishai, A., Ungerleider, L. G., and Haxby, J. V. (2000). Distributed neural systems for the generation of visual images. *Neuron* 28, 979–990. doi: 10.1016/S0896-6273(00)00168-9

Ishida, H., Suzuki, K., and Grandi, L. C. (2014). Predictive coding accounts of shared representations in parieto-insular networks. *Neuropsychologia* 70, 442–454. doi: 10.1016/j.neuropsychologia.2014.10.020

Kanwisher, N., Stanley, D., and Harris, A. (1999). The fusiform face area is selective for faces not animals. *Neuroimage* 10, 183–187. doi: 10.1097/00001756-199901180-00035

Kilner, J. M., Friston, K. J., and Frith, C. D. (2007a). The mirror-neuron system: a Bayesian perspective. *Neuroimage* 18, 619–623. doi: 10.1097/WNR.0b013e3281139ed0

Kilner, J. M., Friston, K. J., and Frith, C. D. (2007b). Predictive coding: an account of the mirror neuron system. *Cogn. Process.* 8, 159–166. doi: 10.1007/s10339-007-0170-2

Kilner, J., Hamilton, A. F., and Blakemore, S. J. (2007c). Interference effect of observed human movement on action is due to velocity profile of biological motion. *Soc. Neurosci.* 2, 158–166. doi: 10.1080/17470910701428190

Kilner, J., Paulignan, Y., and Blakemore, S. (2003). An interference effect of observed biological movement on action. *Curr. Biol.* 13, 522–525. doi: 10.1016/S0960-9822(03)00165-9

Klasen, M., Kenworthy, C. N., Mathiak, K. A., Kircher, T. T. J., and Mathiak, K. (2011). Supramodal representation of emotions. *J. Neurosci.* 31, 13635–13643. doi: 10.1523/JNEUROSCI.2833-11.2011

Lahnakoski, J. M., Glerean, E., Salmi, J., Jääskeläinen, I. P., Sams, M., Hari, R., et al. (2012). Naturalistic fMRI mapping reveals superior temporal sulcus as the hub for the distributed brain network for social perception. *Front. Hum. Neurosci.* 6:233. doi: 10.3389/fnhum.2012.00233

Liberman, A. M. (1996). *Speech: A Special Code*. Cambridge, MA: MIT Press.

Liberman, A. M., Hariris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* 54, 358–368. doi: 10.1037/h0044417

Likowski, K. U., Mühlberger, A., Gerdes, A. B., Wieser, M. J., Pauli, P., and Weyers, P. (2012). Facial mimicry and the mirror neuron system: simultaneous acquisition of facial electromyography and functional magnetic resonance imaging. *Front. Hum. Neurosci.* 6:214. doi: 10.3389/fnhum.2012.00214

Llobera, J., Sanchez-Vives, M. V., and Slater, M. (2013). The relationship between virtual body ownership and temperature sensitivity. *J. R. Soc. Interface* 10, 20130300. doi: 10.1098/rsif.2013.0300

Looser, C. E., and Wheatley, T. (2010). The tipping point of animacy: how, when and where we perceive life in a face. *Psychol. Sci.* 21, 1854–1862. doi: 10.1177/0956797610388044

Lucas, G. M., Gratch, J., King, A., and Morency, L.-P. (2014). It's only a computer: virtual humans increase willingness to disclose. *Comput. Hum. Behav.* 37, 94–100. doi: 10.1016/j.chb.2014.04.043

MacDorman, K. F. (2005). Androids as an experimental apparatus: Why is there an uncanny valley and can we exploit it? *CogSci-2005 Workshop: Towards Social Mechanisms of Android Science*, 106–118.

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

Martini, M., Perez-Marcos, D., and Sanchez-Vives, M. V. (2014). Modulation of pain threshold by virtual body ownership. *Eur. J. Pain* 18, 1040–1048. doi: 10.1002/j.1532-2149.2014.00451.x

McDonnell, R., Breidt, M., and Bülthoff, H. H. (2012). Render me real? Investigating the effect of render style on the perception of animated virtual humans. *ACM Trans. Graph.* 31:91. doi: 10.1145/2185520.2185587

Mellet, E., Tzourio, N., Crivello, F., Joliot, M., Denis, M., and Mazoyer, B. (1996). Functional anatomy of spatial mental imagery generated from verbal instructions. *J. Neurosci.* 16, 6504–6512.

Milgram, S. (1963). Behavioral study of obedience. *J. Abnorm. Soc. Psychol.* 67, 371–378. doi: 10.1037/h0040525

Mitchell, W. J., Szerszen, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., and Macdorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception* 2, 10–12. doi: 10.1068/i0415

Moore, R. K. (2012). A Bayesian explanation of the 'uncanny valley' effect and related psychological phenomena. *Sci. Rep.* 2: 864. doi: 10.1038/srep00864

Mori, M. (1970). The uncanny valley. *Energy* 7, 33–35. [Republished in *IEEE Robotics and Automation Magazine*, June 2012, 98–100].

Moser, E., Derntl, B., Robinson, S., Fink, B., Gur, R. C., and Grammer, K. (2007). Amygdala activation at 3T in response to human and avatar facial expressions of emotions. *J. Neurosci. Methods* 161, 126–133. doi: 10.1016/j.jneumeth.2006.10.016

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., and Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Curr. Biol.* 21, 1641–1646. doi: 10.1016/j.cub.2011.08.031

O'Craven, K. M., and Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J. Cogn. Neurosci.* 12, 1013–1023. doi: 10.1162/08989290051137549

Oztop, E., Franklin, D. W., Chaminade, T., and Cheng, G. (2005). Human-humanoid interaction: is a humanoid robot perceived as a human? *Int. J. HR* 2, 537–559. doi: 10.1142/S0219843605000582

Peck, T. C., Seinfeld, S., Aglioti, S. M., and Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Conscious. Cogn.* 22, 779–787. doi: 10.1016/j.concog.2013.04.016

Piwek, L., McKay, L. S., and Pollick, F. E. (2014). Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition* 130, 271–277. doi: 10.1016/j.cognition.2013.11.001

Press, C., Bird, G., Flach, R., and Heyes, C. (2005). Robotic movement elicits automatic imitation. *Cogn. Brain Res.* 25, 632–640. doi: 10.1016/j.cogbrainres.2005.08.020

Press, C., Gillmeister, H., and Heyes, C. (2007). Sensorimotor experience enhances automatic imitation of robotic action. *Proc. R. Soc. B* 274, 2509–2514. doi: 10.1098/rspb.2007.0774

Rizzo, A. A., Neumann, U., Enciso, R., Fidaleo, D., and Noh, J. Y. (2012). Performance-driven facial animation: basic research on human judgments of emotional state in facial avatars. *Cyberpsychol. Behav.* 4, 471–487. doi: 10.1089/109493101750527033

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2, 661–670. doi: 10.1038/35090060

Sanchez-Vives, M. V., and Slater, M. (2005). From presence to consciousness through virtual reality. *Nat. Rev. Neurosci.* 6, 332–339. doi: 10.1038/nrn1651

Sarkheil, P., Goebel, R., Schneider, F., and Mathiak, K. (2013). Emotion unfolded by motion: a role for parietal lobe in decoding dynamic facial expressions. *Soc Cogn. Affect. Neurosci.* 8, 950–995. doi: 10.1093/scan/nss092

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., and Frith, C. (2011). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* 7, 413–422. doi: 10.1093/scan/nsr025

Sciutti, A., Bisio, A., Nori, F., Metta, G., Fadiga, L., Pozzo, T., et al. (2012). Measuring human-robot interaction through motor resonance. *Int. J. Soc. Robot.* 4, 223–234.

Sel, A. (2014). Predictive codes of interoception, emotion, and the self. *Front. Psychol.* 5:189. doi: 10.3389/fpsyg.2014.00189

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends Cogn. Sci.* 17, 565–573. doi: 10.1016/j.tics.2013.09.007

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., et al. (2006). A virtual reprise of the Stanley Milgram obedience experiments. *PLoS ONE* 1:e39. doi: 10.1371/journal.pone.0000039

Slater, M., Lotto, B., Arnold, M. M., and Sanchez-Vives, M. V. (2009). How we experience immersive virtual environments: the concept of presence and its measurements. *Annu. Psicol.* 40, 193–210.

Slater, M., Rovira, A., Southern, R., Swapp, D., Zhang, J. J., Campbell, C., et al. (2013). Bystander responses to a violent incident in an immersive virtual environment. *PLoS ONE* 8:e52766. doi: 10.1371/journal.pone.0052766

Slater, M., Spanlang, B., and Corominas, D. (2010). Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *ACM Trans. Graph.* 29:92. doi: 10.1145/1778765.1778829

Stienen, B. M. C., Tanaka, A., and de Gelder, B. (2011). Emotional voice and emotional body postures influence each other independently of visual awareness. *PLoS ONE* 6:e25517. doi: 10.1371/journal.pone.0025517

Tai, Y. F., Scherfler, C., Brooks, D. J., Sawamoto, N., and Castiello, U. (2004). The human premotor cortex is 'mirror' only for biological actions. *Curr. Biol.* 14, 117–120. doi: 10.1016/j.cub.2004.01.005

Tapus, A., Tapus, C., and Mataric, M. J. (2008). User-robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intell. Serv. Robot.* 1, 169–183. doi: 10.1007/s11370-008-0017-4

Tay, B., Jung, Y., Park, T. (2014). When stereotypes meet robots: the double-edge sword of robot gender and personality stereotypes in human-robot interaction. *Comput. Hum. Behav.* 38, 75–84. doi: 10.1016/j.chb.2014.05.014

Thompson, J. C., Trafton, J. G., and McKnight, P. (2011). The perception of humanness from the movements of synthetic agents. *Perception* 40, 695–704. doi: 10.1068/p6900

Tikka, P., Väljamäe, A., de Borst, A. W., Pugliese, R., Ravaja, N., Kaipainen, M., et al. (2012). Enactive cinema paves way for understanding complex real-time social interaction in neuroimaging experiments. *Front. Hum. Neurosci.* 6:298. doi: 10.3389/fnhum.2012.00298

Travers, P. (2001). *Final Fantasy*. Available at: http://www.rollingstone.com/movies/reviews/final-fantasy-20010706 [accessed August 1, 2014].

Trojano, L., Grossi, D., Linden, D. E., Formisano, E., Hacker, H., Zanella, F. E., et al. (2000). Matching two imagined clocks: the functional anatomy of spatial analysis in the absence of visual stimulation. *Cereb. Cortex* 10, 473–481. doi: 10.1093/cercor/10.5.473

Vander Wyck, B. C., Hudac, C. M., Carter, E. J., Sobel, D. M., and Pelphrey, K. A. (2009). Action understanding in the superior temporal sulcus region. *Psychol. Sci.* 20, 771–777. doi: 10.1111/j.1467-9280.2009.02359.x

Watson, R., and de Gelder, B. (2014). *Investigating Implicit Crossmodal Decoding of Body-Voice Emotion Using Multivoxel Pattern Analysis*. Program No. 724.15/HH6, 2014 Neuroscience Meeting Planner. Washington, DC: Society for Neuroscience.

Waugh, R. (2012). *Living doll? 'Geminoid F' is most convincing 'robot woman' ever - she has 65 facial expressions, talks and even sings*. Available at: http://www.dailymail.co.uk/sciencetech/article-2128115/Living-doll-Geminoid-F-convincing-robot-woman–facial-expressions-talks-sings.html [accessed August 1, 2014].

Weyers, P., Muhlberger, A., Hefele, C., and Pauli, P. (2006). Electromyographic responses to static and dynamic avatar emotional facial expressions. *Psychophysiology* 43, 450–453. doi: 10.1111/j.1469-8986.2006.00451.x

Weyers, P., Muhlberger, A., Kund, A., Hess, U., and Pauli, P. (2009). Modulation of facial reactions to avatar emotional faces by nonconscious competition priming. *Psychophysiology* 46, 328–335. doi: 10.1111/j.1469-8986.2008.00771.x

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., and Rizzolatti, G. (2003). Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. *Neuron* 40, 655–664. doi: 10.1016/S0896-6273(03)00679-2

Yamada, Y., Kawabe, T., and Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *Jpn. Psychol. Res.* 55, 20–32. doi: 10.1111/j.1468-5884.2012.00538.x

Zucker, N. L., Green, S., Morris, J. P., Kragel, P., Pelphrey, K. A., Bulik, C. M., et al. (2011). Hemodynamic signals of mixed messages during a social exchange. *Neuroreport* 22, 413–418. doi: 10.1097/WNR.0b013e3283455c23

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

REVIEW
published: 10 April 2015
doi: 10.3389/fpsyg.2015.00390

A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness

*Jari Kätsyri *, Klaus Förger, Meeri Mäkäräinen and Tapio Takala*

Department of Computer Science, School of Science, Aalto University, Espoo, Finland

The uncanny valley hypothesis, proposed already in the 1970s, suggests that almost but not fully humanlike artificial characters will trigger a profound sense of unease. This hypothesis has become widely acknowledged both in the popular media and scientific research. Surprisingly, empirical evidence for the hypothesis has remained inconsistent. In the present article, we reinterpret the original uncanny valley hypothesis and review empirical evidence for different theoretically motivated uncanny valley hypotheses. The uncanny valley could be understood as the naïve claim that any kind of human-likeness manipulation will lead to experienced negative affinity at close-to-realistic levels. More recent hypotheses have suggested that the uncanny valley would be caused by artificial–human categorization difficulty or by a perceptual mismatch between artificial and human features. Original formulation also suggested that movement would modulate the uncanny valley. The reviewed empirical literature failed to provide consistent support for the naïve uncanny valley hypothesis or the modulatory effects of movement. Results on the categorization difficulty hypothesis were still too scarce to allow drawing firm conclusions. In contrast, good support was found for the perceptual mismatch hypothesis. Taken together, the present review findings suggest that the uncanny valley exists only under specific conditions. More research is still needed to pinpoint the exact conditions under which the uncanny valley phenomenon manifests itself.

Keywords: uncanny valley, human-likeness, anthropomorphism, perceptual mismatch, categorical perception, computer animation

OPEN ACCESS

Edited by:
Marcus Cheetham, University of Zürich, Switzerland

Reviewed by:
Ian Clara, University of Manitoba, Canada
Tyler John Burleigh, University of Guelph, Canada

****Correspondence:***
Jari Kätsyri, Department of Computer Science, School of Science, Aalto University, PO Box 15500, FIN-00076 Aalto, Espoo, Finland
jari.katsyri@aalto.fi

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *15 August 2014*
Accepted: *19 March 2015*
Published: *10 April 2015*

Citation:
Kätsyri J, Förger K, Mäkäräinen M and Takala T (2015) A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. Front. Psychol. 6:390. doi: 10.3389/fpsyg.2015.00390

Introduction

Masahito Mori predicted already in the 1970s that although people would in general have favorable reactions toward increasingly humanlike robots, almost but not fully human robots would be unsettling (Mori, 1970). Mori used a hypothetical curve to characterize this relationship, and coined the sudden dip in this curve at almost humanlike levels as the uncanny valley (**Figure 1**). Although Mori focused on robots and other mechanical devices, the hypothesis was general enough to incorporate other domains as well. Some relevant technological innovations, such as prosthetic limbs and prototypes of anthropomorphic robots, already existed at the time when the uncanny valley hypothesis was published (cf. Mori, 1970). However, the uncanny valley hypothesis has become fully topical only during the last two decades or so, during which computer animation technologies have seen rapid advances. Although highly realistic computer-animated faces can

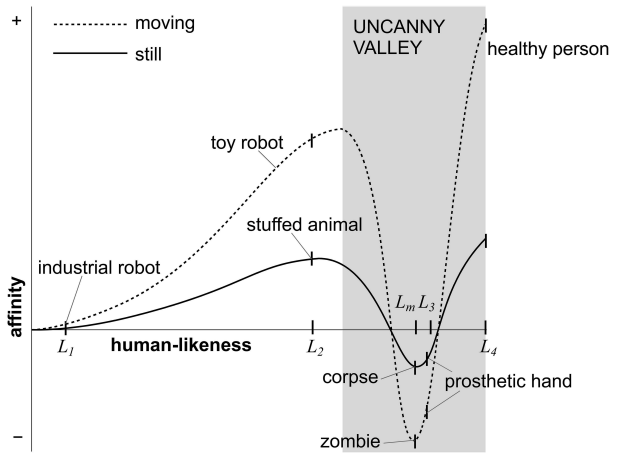


FIGURE 1 | Mori's uncanny valley curve demonstrating the non-linear relationship between the human-likeness of stimuli (clearly artificial to fully human-like) and the observers' sense of affinity for them (negative to positive). Human-likeness levels (*L_1 to L4*) that correspond roughly with the turning points of the curves have been highlighted on the horizontal axis, and the uncanny valley proper has been emphasized with a dark gray color. Adapted with permission from MacDorman (2005).

already be produced (e.g., Alexander et al., 2010; Perry, 2014), contemporary computer animation techniques still tend to suffer from subtle imperfections related for example to rendering, lighting, surface materials, and movement dynamics. Hence, it is not surprising that the uncanny valley hypothesis has been adopted to explain the poor commercial success of some animated films in the media (cf. citations in Brenton et al., 2005; Geller, 2008; Eberle, 2009; Misselhorn, 2009; Pollick, 2010). The uncanny valley hypothesis has also motivated research in various fields beyond robotics and computer animation including, but not limited to, developmental psychology (Matsuda et al., 2012), neuroimaging (e.g., Cheetham et al., 2011; Saygin et al., 2012), animal studies (Steckenfinger and Ghazanfar, 2009), Bayesian statistics (Moore, 2012), and philosophy (Misselhorn, 2009).

Against this background, it is surprising that empirical evidence for the uncanny valley hypothesis is still ambiguous if not non-existent. Early research reviews from year 2005 noted the lack of empirical studies on the uncanny valley (Brenton et al., 2005; Gee et al., 2005; Hanson, 2005). To our knowledge, empirical evidence for the existence of the uncanny valley has still not been reviewed systematically. Several reviews have elaborated the original hypothesis and its underlying mechanisms (Ishiguro, 2006, 2007; Tondu and Bardou, 2011) or applied the original hypothesis in specific contexts (Eberle, 2009), but these reviews have not taken clear sides on the existence of the uncanny valley. Two recent reviews have concluded that the empirical evidence for the uncanny valley is either absent or inconsistent (Pollick, 2010; Zlotowski et al., 2013). These reviews have, however, cited direct evidence from relatively few studies that pertained directly to their specific fields (psychology and human-robot interaction, respectively). A possible reason for the lack of empirical research reviews could be that although a plethora of uncanny valley articles have been published, it is difficult to identify which of them have tested the original hypothesis directly and which have been merely derived from it.

It is also possible that there exist not one but many plausible uncanny valley hypotheses. Because the original uncanny valley hypothesis was intended as a broadly applicable guideline rather than an explicit experimental hypothesis (cf. Pollick, 2010), it is likely to be consistent with several more specific hypotheses. Some of these hypotheses could be derived from established psychological constructs and theories. In some cases, minor adjustments to the original uncanny valley hypothesis could be justified. Because the two major dimensions of the uncanny valley—the human-likeness of stimuli and the observers' experience of affinity for them—were not defined clearly in the original uncanny valley formulation, these dimensions could be operationalized in various different ways. Consequently, different uncanny valley studies could end up addressing different theoretical constructs and hypotheses depending on their specific methodological decisions. Because the human-likeness is difficult to operationalize, confounding factors and other alternative explanations could also limit the conclusions that can be drawn from individual studies.

The main goal of the present article was to review up-to-date empirical research evidence for a framework of plausible uncanny valley hypotheses derived from the original uncanny valley article (Mori, 1970) and other more recent publications. The review consists of five major sections. First, we will provide an interpretation of the original human-likeness and affinity dimensions of the uncanny valley (Section An Interpretation of the Uncanny Valley). We will argue that a literal interpretation of Mori's original examples, especially those involving morbid characters (i.e., corpses and zombies), would confound human-likeness with extraneous factors. We will also suggest that the original formulation of the affinity dimension could be interpreted both in terms of perceptual familiarity and emotional valence. Second, we will formulate a framework of empirically testable uncanny valley hypotheses based on the preceding analysis (Section A Framework of Uncanny Valley Hypotheses). In addition, we will reiterate the recent categorization ambiguity and perceptual mismatch hypotheses (e.g., Brenton et al., 2005; Pollick, 2010; Cheetham et al., 2011). Third, we will formulate explicit criteria for article inclusion and evaluation (Section Article Selection and Evaluation). Fourth, we will review empirical evidence for the formulated hypotheses based on the adopted evaluation criteria (Section Review of Empirical Evidence). Finally, we will discuss the implications and limitations of our findings and consider open questions in uncanny valley research (Section Discussion).

An Interpretation of the Uncanny Valley

What Is Human-Likeness?

Human-likeness is not a single quality of artificial characters that could be traced back to specific static, dynamic, or behavioral features—instead, human-likeness could be varied in an almost infinite number of different ways. Mori (1970) himself used anecdotal examples to characterize different degrees

of human-likeness. We have highlighted some of these examples in **Figure 1** and summarized them in **Table 1**. The hypothetical human-likeness levels corresponding with the selected examples have been labeled from L_1 to L_4. Mori used industrial robots (L_1) as an example of the least humanlike characters with any resemblance to real humans. Although clearly artificial, such characters have some remotely humanlike characteristics, such as arms for gripping objects. Stuffed animals and toy robots (L_2) were placed close to the first peak of the uncanny curve. Like industrial robots, these characters are clearly artificial; however, unlike industrial robots, such characters have also been purposefully designed to resemble humans. Mori placed two different kinds of objects or characters near the bottom of the valley. First, Mori mentioned prosthetic hands (L_3) as an example of manmade artifacts that have been meant to appear humanlike but that have failed to do so because of some artificial qualities. Second, Mori mentioned human corpses and zombies (L_m) when considering danger avoidance as a speculative explanation of the uncanny valley. Finally, Mori used healthy humans (L_4) as an example of full human-likeness. In these examples, Mori referred to both static and moving instances of similar characters (e.g., still and animate corpses) to illustrate how movement would amplify the uncanny curve (**Figure 1**).

Table 1 also illustrates two extraneous factors that could affect affinity responses to the above anecdotal examples if they were taken literally. First, stuffed animals and toy robots could elicit positive reactions not only because they appear somewhat humanlike but because they have been purposefully designed to appear aesthetic. Similarly, human corpses, whether still or animate, would certainly not evoke negative reactions only because they appear humanlike but because they are morbid and horrifying. These considerations strongly suggest that Mori's original examples should not be adopted literally in empirical studies. However, once this approach is rejected, the question still remains which human-likeness manipulations should be used in empirical studies out of all imaginable possibilities. Although this question does not yet have an agreed upon answer, there seems to be a trend toward using image morphing and computer graphics (CG) techniques for manipulating facial stimuli in recent studies (cf. Table S1).

TABLE 1 | Focal points on the human-likeness dimension of the uncanny valley graph.

HL	Anecdotal examples	Human-likeness	Extraneous factors	Affinity
L_1	Industrial robot	Clearly artificial	–	Neutral
L_2	Stuffed animal, toy robot	Somewhat humanlike	Aesthetics	Positive
L_3	Prosthetic hand	Almost humanlike	–	Negative
L_m	Corpse, zombie	Almost humanlike	Morbidity	Negative
L_4	Healthy human	Fully humanlike	–	Very positive

Anecdotal examples refer to Mori (1970).
HL—degree of human-likeness.

What Is Affinity?

Mori's original Japanese terms *bukimi* and *shinwakan* (or *shinwakan*) for the affinity dimension referred to several different concepts. The negative term *bukimi* translates quite unequivocally into eeriness (Ho and MacDorman, 2010), although other similar terms such as creepiness and strangeness have also been used (cf. Ho et al., 2008). In contrast, the positive term *shinwakan* is an unconventional Japanese word, which does not have a direct equivalent in English (Bartneck et al., 2007, 2009). The earliest and the most common translation of this term has been familiarity; however, it has been argued that likability would be a more appropriate translation (ibid.). In the latest English translation of Mori's original article, *shinwakan* was translated as affinity (Mori, 1970/2012). Similarly, we have adopted affinity when referring to the *bukimi–shinwakan* dimension in the present article. **Table 2** lists dictionary definitions (Merriam-Webster Online Dictionary; http://www.merriam-webster.com; accessed 24.11.2014) for the most commonly used affinity terms. A closer inspection of these terms would suggest that all of them refer to various aspects of perceptual familiarity and emotional valence. Perceptual familiarity refers to recognizing that the perceived character has similar qualities as another object the observer is already well acquainted with (possibly, the observer himself or herself). Emotional valence covers various positive (liking, pleasantness, and attraction) and negative (aversive sensations) emotions elicited by the character. Although positive and negative affinity could be considered separately (e.g., Ho and MacDorman, 2010), emotional valence is an established psychological concept (e.g., Russell, 2003) that is able to incorporate both of them.

Given that the original terms for the affinity dimension (or at least their common translations) are ambiguous, empirical studies would be necessary for resolving which self-report items would be ideal for measuring affinity. Previous studies have suggested that eeriness is associated with other negative emotion terms such as fear, disgust, and nervousness (Ho et al., 2008); or fear, unattractiveness, and disgust (Burleigh et al., 2013). To our

TABLE 2 | Dictionary definitions for the common English translations of Mori's affinity dimension.

English term	Definitions
Eeriness	[The quality of being] strange and mysterious
	[...] so mysterious, strange, or unexpected as to send a chill up the spine
Likability	[...] easy to like
	[...] pleasant or appealing
	[…bringing] about a favorable regard
Familiarity	The state of being [well acquainted] with something
	[...] having knowledge about something
	A state of close relationship [similar to intimacy]
Affinity	A feeling of closeness and understanding that someone has for another person because of their similar qualities, ideas, or interests
	A liking for or an attraction to something
	The state of being similar or the same

knowledge, only one previous study up to date has used factor analytic methods to develop a conclusive self-report questionnaire for uncanny valley studies (Ho and MacDorman, 2010). This study identified orthogonal factors for human-likeness, eeriness (two separate factors: eerie and spine-tingling), and attractiveness. An informal evaluation would suggest some potential problems with this questionnaire, however. First, some of the questionnaire items are not necessarily ideal for measuring their intended constructs in all contexts. For example, the semantic differential items "ordinary—supernatural" and "without definite lifespan—mortal" could be inappropriate human-likeness measures when none of the evaluated stimuli are supernatural. Second, although the identified eeriness factors are consistent with Mori's original terms, their constituent items (e.g., "numbing—freaky" and "unemotional—hair-rising") do not resemble items in typical emotion self-report questionnaires (cf. self-report items in Bradley and Lang, 1994). Third, familiarity items were not considered in the study, although familiarity would seem to be an integral part of the uncanny valley. Although future empirical studies might be useful for refining this scale, this work is an important step toward developing a common metric for the affinity dimension. The scale has already been applied in at least two studies (Mitchell et al., 2011; MacDorman et al., 2013).

A Framework of Uncanny Valley Hypotheses

Figure 2 illustrates the preceding analysis of the uncanny valley phenomenon (Section An Interpretation of the Uncanny Valley) and the relations between the present hypotheses and the uncanny valley concepts.

Naïve Hypotheses

The question of which specific human-likeness manipulations should be used in empirical uncanny valley studies could be sidestepped by assuming that *any* kind of manipulation would lead to the characteristic uncanny curve for affinity (**Figure 1**). However, this hypothesis is simplistic because it assumes that all imaginable human-likeness manipulations are equally relevant for the uncanny valley. Consequently, it could be referred to as a naïve uncanny valley hypothesis as opposed to more specific hypotheses (Section Refined Hypotheses). We have attempted to formulate this hypothesis so that it would be compatible with various human-likeness manipulations ranging from categorical manipulations with a minimal number of human-likeness levels to fully continuous manipulations. **Figure 3** illustrates the original uncanny curve for the four most focal human-likeness levels (**Table 1**). These levels constitute the minimal set of human-likeness levels that could be used to capture the most relevant aspects of the original uncanny curve.

The core claim of the uncanny valley is that almost humanlike characters will elicit more negative affinity than any other characters (**Figure 3**). As can be seen in the darkened region of **Figure 1**, this characteristic U-shaped curve forms the uncanny valley proper. Because almost humanlike characters would need to be compared to both more artificial and more humanlike characters, the bare minimum for testing this prediction would be three human-likeness levels (cf. **Figure 3**). Although not equally critical, the original uncanny valley hypothesis also predicts that, except for the uncanny valley proper, affinity will be more positive for increasingly humanlike characters. That is, affinity increases when moving from clearly artificial to somewhat humanlike characters, and there would also be a relative increase between somewhat and fully humanlike characters (**Figure 3**). Given that this hypothesis omits almost humanlike characters, at least the remaining three levels in **Figure 3** would need to be used to test this prediction. These predictions can be formulated as the following hypotheses.

H1a ("naïve uncanny valley proper"): For any kind of human-likeness manipulation, almost humanlike characters will elicit more negative affinity (lower familiarity and/or more negative emotional valence) than any other more artificial or more humanlike characters.

H1b ("naïve human-likeness"): For any kind of human-likeness manipulation, more humanlike characters will elicit more positive affinity (higher familiarity and/or more positive emotional valence), with the possible exception of characters fulfilling H1a.

Morbidity Hypothesis

Although purposefully morbid characters could be adopted from the original uncanny valley formulation (Mori, 1970) and used in empirical uncanny valley studies, such characters would confound the more interesting effects of varying human-likeness (Section What is Human-Likeness?). Although it is quite trivial that such characters should evoke negative affinity, we have nevertheless formulated the following hypothesis to help separate morbidity effects from those of other hypotheses.

H1c ("morbidity"): Morbid characters (e.g., corpses or zombies) will elicit more negative affinity (lower familiarity and/or more negative emotional valence) than any other characters.

Movement Hypotheses

In his original formulation, Mori (1970) also suggested that movement would amplify the uncanny curve. That is, the positive and negative affinity experiences elicited by the still characters should become more pronounced for moving characters. The role of movement could, however, be more complex than originally predicted. For example, although Mori considered movement as a dichotomous variable—it either is or is not present—movement features could also range in human-likeness and lead to an uncanny curve of their own. This leads to the following reformulations of the naïve uncanny hypotheses (H1a and H1b).

H2a ("uncanny valley proper for movement"): For any kind of human-likeness manipulation, "almost humanlike" movement patterns will elicit more negative affinity (lower familiarity and/or more negative emotional valence) than any other more artificial or more humanlike movement patterns.

H2b ("human-likeness for movement"): For any kind of human-likeness manipulation, more humanlike movement patterns will elicit more positive affinity (higher familiarity and/or more positive

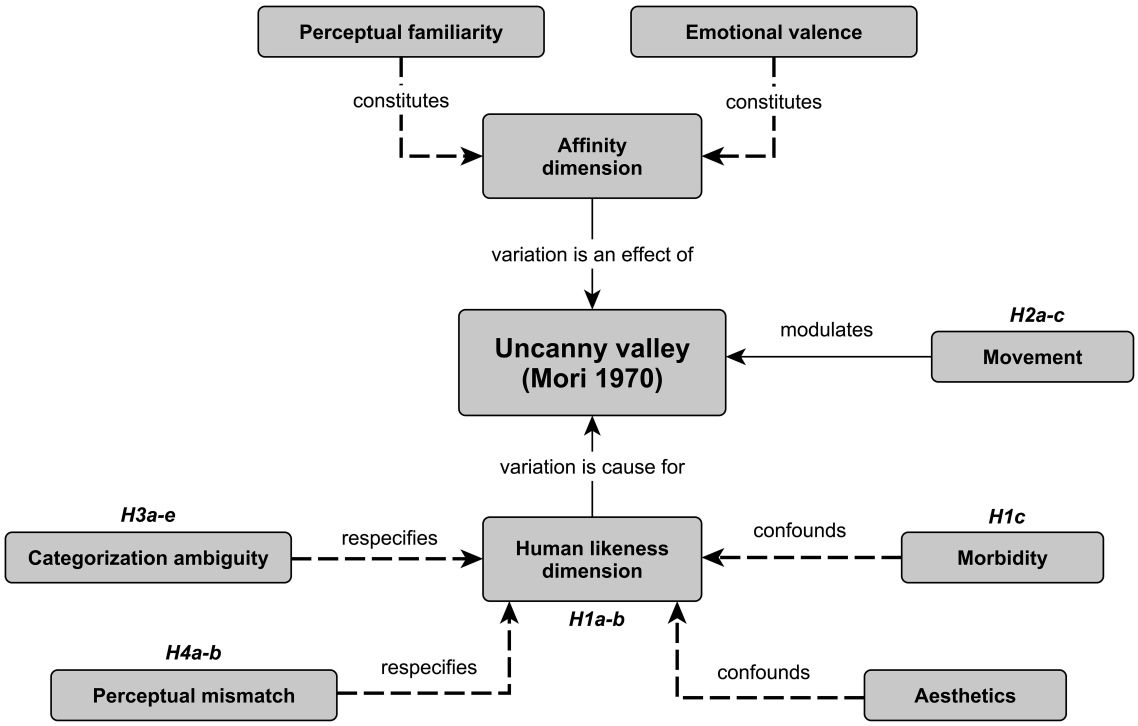


FIGURE 2 | A concept map demonstrating relations between the present uncanny valley hypotheses and different uncanny valley concepts derived from Mori (1970). Dashed lines refer to constructs that have been explicated after Mori's original publication. Hypotheses: H1a—naïve UV proper, H1b—naïve HL, H1c—morbidity, H2a—UV proper for movement, H2b—HL for movement, H2c—movement modulation, H3a—category identification, H3b—perceptual discrimination, H3c—categorical identification difficulty, H3d—opposite perceptual discrimination, H3e—perceptual discrimination difficulty, H4a—inconsistent HL, H4b—atypicality; UV—uncanny valley, HL—human-likeness.

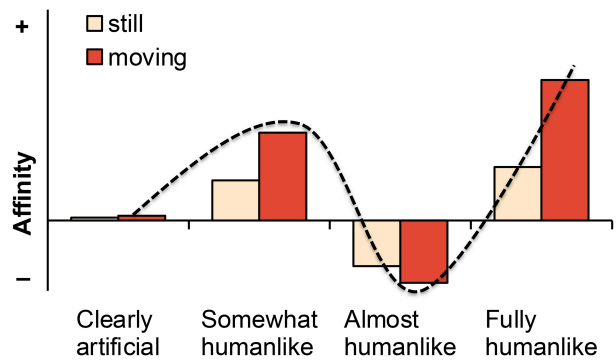


FIGURE 3 | Predicted affinity levels (from negative to positive) for still and moving versions of characters representing different human-likeness levels. The characteristic uncanny curve is overlain on the data for illustration.

emotional valence), with the possible exception of movement patterns fulfilling H2a

The original movement hypothesis can be stated as follows.

H2c ("movement modulation"): Movement will amplify the affinity responses (changes in familiarity and/or emotional valence) associated with hypotheses H1a and H1b.

Testing movement hypotheses H2a–c would require the same number of minimum human-likeness levels as the more general hypotheses H1a and H1b (that is, three levels; the specific levels depending on the hypothesis).

Refined Hypotheses

Categorization Ambiguity

Early uncanny valley postulations have suggested that negative affinity would be caused by the ambiguity in categorizing highly realistic artificial characters as real humans or artificial entities (e.g., Ramey, 2005 [quoted in MacDorman and Ishiguro, 2006]; Pollick, 2010). Notably, this suggestion itself does not yet consider whether the human-likeness dimension of the uncanny valley is perceived continuously or categorically—that is, some intermediate characters could be difficult to categorize regardless of whether increasing human-likeness were perceived as a gradual continuum or discretely as artificial and human categories.

Categorical perception, which is an empirically and theoretically established construct in psychology, has been applied to the uncanny valley in recent empirical studies (Cheetham et al., 2011, 2014). Loosely speaking, categorical perception refers to the phenomenon where the categories possessed by an observer influence his or her perceptions (Goldstone and Hendrickson, 2010). Specifically, categorical perception is thought to occur

when the perceptual discrimination is enhanced for pairs of perceptually adjacent stimuli straddling a hypothetical category boundary between two categories, and decreased for equally spaced pairs belonging to the same category (Repp, 1984; Harnad, 1987; Goldstone and Hendrickson, 2010). Applied to the uncanny valley, categorical perception would mean that "[…] irrespective of physical differences in humanlike appearance, objects along the DOH [degree of human-likeness] are treated as conceptually equivalent members of either the category 'non-human' or the category 'human,' except at those levels of physical realism at the boundary between these two categories." (Cheetham et al., 2011, p. 2).

The two most commonly agreed upon criteria for experimental demonstrations of categorical perception are category identification and perceptual discrimination (Repp, 1984; Harnad, 1987). The identification criterion means that stimulus identification in a labeling task should follow a steep slope such that labeling probabilities change abruptly at the hypothetical category boundary. Given that the location of category boundary cannot be known in advance, the minimum number of required stimulus levels for testing this hypothesis cannot be determined precisely. In practice, previous uncanny valley studies have employed at least 11 evenly distributed human-likeness steps along the human-likeness continuum (e.g., Looser and Wheatley, 2010; Cheetham et al., 2011, 2014). Response times have been used as an index of uncertainty in the identification task (e.g., Pisoni and Tash, 1974; de Gelder et al., 1997). Assuming that categorization ambiguity should be the greatest at the category boundary, the slowest response times should also coincide with this point. That is,

H3a ("category identification"): A steep category boundary will exist on the human-likeness axis such that characters on the left and right sides of this boundary are labeled consistently as "artificial" and "human," respectively; and/or this identification task will elicit the slowest response times at the category boundary.

The discrimination criterion refers to the above requirement that perceptual discrimination should be better for stimulus pairs straddling the category boundary than for equally spaced stimulus pairs falling on the same side of the category boundary. As an example from previous work (Cheetham et al., 2011), the four stimulus pairs artificial–artificial, artificial–human, human–artificial, and human–human could be derived from identification results and employed in the perceptual discrimination task. All possible stimulus pairs that are differentiated by an equal number of steps on the human-likeness continuum could also be used (e.g., Cheetham et al., 2014). As a summary, to demonstrate categorical perception for the human-likeness dimension of the uncanny valley, the following hypothesis should be confirmed in addition H3a.

H3b ("perceptual discrimination"): Character pairs that straddle the category boundary between "artificial" and "human" categories will be easier to discriminate perceptually than equally different character pairs located on the same side of the boundary.

After demonstrating that the human-likeness dimension is perceived categorically, it would still need to be shown that category identification difficulty (i.e., H3a) is also associated with subjective experiences of negative affinity. The most straightforward assumption would be that identification uncertainty at the category boundary ("categorization ambiguity") leads to negative affinity. Strictly speaking, this hypothesis is not fully consistent with categorical perception as it is commonly understood, given that the hypothesis refers only to the category identification criterion (cf. H3a), whereas perceptual discrimination criterion (H3b) has been considered as the hallmark of categorical perception (e.g., Harnad, 1987). Hence, categorization ambiguity could lead to negative affinity even in the absence of categorical perception (i.e., when only H3a but not H3b holds true). However, we have included this hypothesis, as it is consistent with the early uncanny valley literature (Ramey, 2005 [quoted in MacDorman and Ishiguro, 2006]; Pollick, 2010). Hence, we have formulated this hypothesis as follows.

H3c ("categorical identification difficulty"): Characters that are located at the category boundary between "artificial" and "human" categories (as identified in H3a-b) will elicit more negative affinity (lower familiarity and/or more negative emotional valence) than any other characters that are located on the left or right sides of the category boundary.

As suggested recently by Cheetham et al. (2014), the original uncanny valley hypothesis is based on the implicit assumption that perceptual discrimination is the most difficult for characters in the uncanny valley. However, assuming that the uncanny valley proper is thought of as coinciding with the category boundary and that this boundary is considered in terms of the categorical perception framework, it follows (as in Cheetham et al., 2014) that perceptual discrimination performance should actually be easier for characters at or in the close vicinity of the category boundary and not more difficult. As can be seen, this is the position taken in hypothesis H3b, and the perceptual discrimination difficulty assumption would be its opposite. Assuming that perceptual discrimination would be more difficult for characters in the uncanny valley, this difficulty should also be associated with negative affinity. These hypotheses can be stated as follows.

H3d ("opposite perceptual discrimination"): Character pairs that straddle the category boundary between "artificial" and "human" categories will be more difficult to discriminate perceptually than equally different character pairs located on the same side of the boundary.

H3e ("perceptual discrimination difficulty"): Increased perceptual discrimination difficulty for adjacent character pairs will be associated with heightened negative affinity (lower familiarity and/or more negative emotional valence).

Perceptual Mismatch

Hypotheses H3a-e are attractive because they are related to the well-established framework of categorical perception. However, there are several reasons for considering also other alternatives to these categorization ambiguity and categorical perception

based explanations. First, there is no *a priori* reason for expecting that the human-likeness dimension should be perceived categorically rather than continuously. For example, Campbell et al. (1997) has demonstrated that whereas morphed continua between human and cow faces are perceived categorically, similar continua between humans and monkeys are continuous. Similarly as humans and other primates, humans and anthropomorphic characters share many fundamental similarities that could place them in the same overarching category of humanlike entities (cf. Campbell et al., 1997; Cheetham et al., 2011). Second, negative affinity could of course be caused by some other mechanisms in addition to (or instead of) categorization ambiguity or categorical perception. For example, it is conceivable that some characters on the "human" side of the category boundary would be considered eerie because they appeared human but contained features that are not "entirely right." In this hypothetical but conceivable example, a negative affinity peak would be located on the right side of the category boundary.

The perceptual mismatch hypothesis, which is theoretically independent from the categorization ambiguity and categorical perception hypotheses, has been presented recently as another explanation for the uncanny valley (e.g., MacDorman et al., 2009; Pollick, 2010). This hypothesis suggests that negative affinity would be caused by an inconsistency between the human-likeness levels of specific sensory cues. Clearly artificial eyes on an otherwise fully human-like face—or vice versa—is an example of such inconsistency. A particularly interesting proposal is that negative affinity would be caused by inconsistent static and dynamic information (Brenton et al., 2005; Pollick, 2010). The bare minimum for testing this hypothesis would be four experimental manipulation levels (i.e., two realism levels × two different features). We have formulated this hypothesis in more general terms below.

H4a ("inconsistent human-likeness"): Characters with inconsistent artificial and humanlike features will elicit more negative affinity (lower familiarity and/or more negative emotional valence) than characters with consistently artificial or characters with consistently humanlike features.

Another form of perceptual mismatch could be higher sensitivity to deviations from typical human norms for more humanlike characters (e.g., Brenton et al., 2005; MacDorman et al., 2009). Deviations from human norms could result, for example, from such atypical features as grossly enlarged eyes. In the uncanny valley context, a plausible explanation for this phenomenon could be that the human visual system has acquired more expertise with the featural restrictions of other humans than with the featural restrictions of artificial characters (cf. Seyama and Nagayama, 2007). This hypothesis is also consistent with previous studies demonstrating that faces with typical or average features are considered more attractive than atypical faces (e.g., Langlois and Roggman, 1990; Rhodes et al., 2001). The atypicality hypothesis is similar to the above inconsistency hypothesis, given that atypical features could also be considered artificial. In fact, these two hypotheses have previously been considered as the same hypothesis (e.g., MacDorman et al., 2009). However, the atypicality hypothesis could refer to any deviant features besides artificiality (e.g., any distorted human features) and, unlike the inconsistency hypothesis, it makes a unilateral prediction related to only humanlike characters. Testing atypicality would require at least four experimental manipulation levels (artificial without atypical features, artificial with atypical features, human without atypical features, human with atypical features), and it could be formulated as follows.

H4b ("atypicality"): Humanlike characters with atypical features will elicit more negative affinity (lower familiarity and/or more negative valence) than artificial characters with atypical features, or either humanlike or artificial characters without atypical features.

Relation to the Original Uncanny Valley Hypothesis

The above hypotheses can be seen as refinements of the original uncanny valley hypothesis such that each of them narrows the human-likeness conditions under which the uncanny valley is expected to occur. These hypotheses pertain only to the uncanny valley proper (i.e., the "almost humanlike" level), and they cannot account for the first peak in the uncanny curve (cf. H1b and **Figure 3**). Otherwise, all of these hypotheses would appear to be consistent with the original uncanny valley hypothesis. For example, all of them seem to be consistent with the following quote: "One might say that the prosthetic hand has achieved a degree of resemblance to the human form [. . .]. However, once we realize that the hand that looked real at first sight is actually artificial, we experience an eerie sensation." (Mori, 1970, p. 99; see also MacDorman et al., 2009, p. 698). Here, the prosthetic hand could have appeared eerie because it caused an artificial–human category conflict (H3), it was perceived as containing mismatching artificial and human features (H4a), or because the hand resembled a real hand without fulfilling all of the typical characteristics of human hands (H4b).

Article Selection and Evaluation

Evaluation Criteria

Table 3 displays the criteria that were used for selecting individual studies and for evaluating their results. These criteria are based on the general validity typology of Shadish et al. (2002), which describes four different types of validity and their associated threats. Our goal was to identify justifiable and plausible threats for conclusions that can be drawn from the reviewed studies to hypotheses H1–H4. Hence, we have not attempted to develop a comprehensive list of all possible threats to the experimental validity of individual studies.

Statistical Conclusion Validity

Statistical conclusion validity refers to the validity of inferring that the experimental manipulations and measured outcomes covaried with each other. At the bare minimum, any kind of statistical test should be used to provide evidence against chance results. The predicted U-shaped relationship between human-likeness and affinity (**Figure 1**) could be tested, for example, by using second-order correlation tests or analysis of variance followed by *post-hoc* comparisons. Linear correlation test would, however, not be sufficient for testing the predicted nonlinear

TABLE 3 | Evaluation criteria for possible threats that limit the conclusions that could be drawn from individual studies to the present hypotheses.

Threat	Validity type
No or inadequate statistical tests[a]	Statistical conclusion
Heterogeneous stimuli	Statistical conclusion
No manipulation check for human-likeness[a]	Internal
Image morphing artifacts	Internal
Categorical perception not tested[b]	Construct
Irrelevant affinity measures[a]	Construct
Familiarity evaluations misunderstood	Construct
Outlier stimuli (e.g., morbid characters)	Construct
Alternative explanations	Construct
Narrow human-likeness range	Construct
Narrow set of manipulated stimuli	Construct
Narrow participant sample	External

Validity types refer to Shadish et al. (2002).
[a] *Used as article inclusion criteria.*
[b] *Applies only to the hypotheses H3c and H3e.*

relationship. Statistical conclusion validity could also be compromised by uncontrolled variation in the stimuli. This issue could be a particular concern for realistic stimuli (e.g., video game characters), whose features cannot be fully controlled. Extraneous variation could possibly be reduced by careful pretesting of stimuli and the inclusion of a large number of stimuli for each stimulus category.

Internal Validity

Internal validity refers to whether the observed outcomes were caused solely by experimental manipulations or whether they would have occurred even without them. Failure to check or confirm that human-likeness manipulations elicited consistent changes in perceived human-likeness would raise doubts over whether human-likeness was actually varied as intended, and would hence threaten internal validity.

Artifacts produced by human-likeness manipulations could also be considered as threats to internal validity (strictly speaking, these and any other confounds would be threats to construct validity in the original typology; cf. Shadish et al., 2002, p. 95). We will consider image morphing artifacts in detail because this method has become popular in uncanny valley studies (cf. Table S1). Image morphing procedure is used to construct a sequence of gradual changes between two images (e.g., CG and human faces), and it consists of three phases: geometric correspondence is established between the images, a warping algorithm is applied to match the shapes of the original objects, and color values are interpolated between the original and warped images (e.g., Wolberg, 1998). Image morphing algorithms are prone to at least two kinds of artifacts (e.g., Wu and Liu, 2013). First, ghosting or double-exposure between images can occur if they contain different features, geometric correspondence has not been established adequately, or warping has not been applied. Second, color interpolation typically causes some blurring because it combines values from several pixels in the original images. Image morphing artifacts are a threat to validity because they are likely to coincide with intermediate levels of human-likeness (i.e., the most processed images). Cheetham and Jäncke (2013) have published a detailed guideline for applying morphing to facial images in uncanny valley studies. We have adopted the following criteria from their guideline: (i) several morphed continua should be used, (ii) selected endpoint images should be similar to each other (i.e., the faces should have similar geometries, have neutral facial expressions, and represent individuals of similar ages), (iii) alignment disparities should be avoided, and (iv) any external features should be masked (i.e., hair and ears, jewelry, and other external features).

Construct Validity

Construct validity refers to the extent to which the experimental manipulations and measured outcomes reflect their intended cause and effect constructs. For example, if the categorization ambiguity hypothesis (H3c or H3e) were demonstrated for specific stimuli without also demonstrating that these stimuli indeed were perceived categorically (H3a–b), it could be uncertain whether categorical perception was in fact involved. For the present purposes, we have required that the outcome measures should tap into the perceptual familiarity and/or emotional valence constructs (Section What is Affinity?). A specific threat related to self-reported familiarity is that it could be confounded with previous experience (e.g., a video game character could be familiar because of its popularity). The inclusion of outlier stimuli that represent other constructs besides varying human-likeness, for example morbidity (Section What is Human-Likeness?), would also threaten construct validity. In the present context, the hypothesis H1c was intended to set such constructs apart from human-likeness. It is also possible that affinity changes could in some cases be explained by other alternative constructs or phenomena (e.g., poor lip synchronization). A narrow range of manipulated human-likeness (e.g., only CG characters) could threaten construct validity because the results would not necessarily generalize to the full range of human-likeness. Application of human-likeness manipulations to only a single stimulus character could also threaten construct validity, if it were plausible that the manipulation results would contain other irrelevant features in addition to or instead of human-likeness.

External Validity

External validity refers to what extent the observed causal relationship between manipulated and observed variables can be generalized to other participants, experimental manipulations, and measured outcomes. Generalizability could be considered by comparing results from different studies. In practice, this would be difficult because of the heterogeneity of uncanny valley studies (cf. Table S1). For the present purposes, we have considered external validity only to exclude results from individual studies with clearly unrepresentative participant samples (e.g., only children).

Article Selection

We identified empirical uncanny valley studies by searching for the key term "uncanny valley" in the following search engines: Scopus (search in article title, abstract, and keywords; including

secondary documents; $N = 273$), PubMed (search in all fields; $N = 23$), Science Direct (search in all fields; $N = 134$), and Web of Science (search in topic; $N = 114$). The obtained list of articles was augmented by other articles cited in them and by articles identified from other sources ($N = 6$). This initial list ($N = 550$) was screened by the first author. Duplicate entries and other than full-length articles published in peer-reviewed journals or conference proceedings were removed semi-automatically, and a cursory selection was done to exclude studies that had clearly not tested or considered the present hypotheses.

The screened list ($N = 125$) was evaluated by all authors for eligibility. The following inclusion criteria were used (cf. **Table 3**): (i) the study had addressed, implicitly or explicitly, at least one of the hypotheses H1–H4; (ii) the study had used at least the minimum number of human-likeness levels for each hypothesis (cf. Section A Framework of Uncanny Valley Hypotheses); (iii) human-likeness of stimuli had been tested explicitly and confirmed; (iv) unless irrelevant for the tested hypothesis (i.e., H3a, H3b, and H3d), the study had used any of the conventional self-report items (likability, eeriness, familiarity, or affinity) or their equivalents for measuring affinity responses; and (v) justified statistical test had been used for testing the relationship between human-likeness and affinity. Two studies that had not tested human-likeness explicitly (Seyama and Nagayama, 2007; Mäkäräinen et al., 2014) were nevertheless included because their human-likeness manipulations (image morphing from artificial to human faces and increasingly more abstract image manipulations, respectively) should have been expected to elicit trivial changes in perceived human-likeness. The final list of selected articles ($N = 17$) is given in Table S1.

Article and Hypothesis Evaluation

The validity of conclusions from individual studies to hypotheses H1–H4 was evaluated using those evaluation criteria in **Table 3** that had not already been adopted as inclusion criteria. All threats that were considered possible are listed in Table S1; however, only those threats that were considered both plausible and relevant for a specific hypothesis were used for excluding individual results. To allow critical evaluation and possible reanalysis of the present findings, we have attempted to highlight potential controversies related to the inclusion and evaluation of studies when reviewing the evidence for each hypothesis.

Because the selected articles had used heterogeneous methodologies and most of them had not reported effect size statistics, a quantitative meta-analysis would not have been appropriate. Instead, we opted to present the numbers of findings providing significant and non-significant evidence for each hypothesis. Because significant findings opposite to hypotheses were rare, they were pooled with the non-significant findings. Significant opposite findings have been mentioned separately in the text. Although this kind of "box score" approach is inferior to quantitative meta-analytic methods (Green and Hall, 1984), it can nevertheless be used to provide an overall quantification of result patterns in the reviewed literature. Following a previous recommendation (Green and Hall, 1984), we adopted a 30% threshold for deciding how many positive findings would be considered significant evidence in favor of a specific hypothesis. All of the reported findings were clearly above this threshold.

Review of Empirical Evidence

Naïve and Morbidity Hypotheses

Empirical evidence for naïve, morbidity, and movement hypotheses is presented in **Table 4**. Whereas the results clearly confirmed that affinity increased linearly across increasing human-likeness (H1b; 7 out of 9 studies), the predicted uncanny valley proper (H1a) received almost no support (1 out of 8 studies). As an exception, one study showed that pictures of intermediate prosthetic hands were more eerie than pictures of either mechanical or human hands (Poliakoff et al., 2013). Two other studies provided results that resembled the uncanny curve (McDonnell et al., 2012; Piwek et al., 2014); however, closer inspection suggested that these results could have been explained by outlier stimuli—that is, in terms of the hypothesis H1c. Another one of these studies (McDonnell et al., 2012) could have provided evidence for H1a even after the outlier stimulus (purposefully ill character) was excluded. However, we considered this evidence inconsistent because both unrealistic ("Toon-Bare" rendering) and realistic ("HumanBasic" rendering) stimuli were found to be less appealing, friendly, and trustworthy than the remaining stimuli.

One of the studies in **Table 4** (Yamada et al., 2013) was excluded from the total count because of plausible morphing artifacts. This study found a U-shaped curve for self-reported pleasantness vs. morphed human-likeness, which could have been

TABLE 4 | Empirical evidence for hypotheses H1 (naïve hypotheses and morbidity) and H2 (movement).

Author/year	H1a	H1b	H1c	H2a	H2b	H2c
Seyama and Nagayama, 2007	–	–				
MacDorman et al., 2009	–	+				
Looser and Wheatley, 2010	–	+				
Thompson et al., 2011				–	+	
McDonnell et al., 2012	(+)	+	+			(+)
Yamada et al., 2013	(+)	(–)				
Burleigh et al., 2013	–	+				
Carter et al., 2013	–	+				
Poliakoff et al., 2013	+	(+)				
Cheetham et al., 2014	–	+				
Piwek et al., 2014	(+)	–	+	–	+	(–)
Rosenthal–von der Pütten and Krämer, 2014	–	+				
Total	8	9	2	2	2	0
+	1	7	2	0	2	0
–	7	2	0	2	0	0

Conclusions: "+": significant in favor of the hypothesis, and "–": non-significant or significant against the hypothesis. Conclusions in parentheses have been omitted from total scores because of plausible threats to validity. Hypotheses: H1a—naïve UV proper, H1b—naïve HL, H1c—morbidity, H2a—UV proper for movement, H2b—HL for movement, H2c—movement modulation. UV—uncanny valley, HL—human-likeness.

taken as support for H1a. However, in this study, only one pair of images had been selected for creating the human-likeness continuum, the selected cartoon and human face were very dissimilar from each other, and no masking had been used (cf. Section Article Selection; and Cheetham and Jäncke, 2013). Hence, it is possible that the lower pleasantness ratings for intermediate morphs could have resulted from morphing artifacts rather than intermediate human-likeness level. Consistently with this interpretation, other morphing studies (Looser and Wheatley, 2010; Cheetham et al., 2014) with masked faces and multiple matched face pairs have failed to find a similar U-shaped curve for participants' evaluations. Another morphing study in **Table 4** (Seyama and Nagayama, 2007) had also used unmasked and quite dissimilar face pairs; however, it is unlikely that the lack of significant findings in this study could have been explained by morphing artifacts.

Several other potentially interesting studies were excluded during the initial selection and were hence not included in **Table 4** or Table S1. For example, seminal uncanny valley studies (Hanson, 2006; MacDorman, 2006; MacDorman and Ishiguro, 2006) were excluded because these studies did not report statistical test results for their findings. Because these studies also seemed to be influenced by morphing artifacts or the use of heterogeneous stimuli, their results for hypotheses H1a–b would nevertheless have been excluded as per our evaluation criteria. Results from several studies using realistic video game (or similar) characters have also been excluded either because they had not used statistical tests or because they had tested only linear correlations statistically. Most of the excluded studies had also deliberately included outlier characters (e.g., zombies) in their experimental stimuli (e.g., Schneider et al., 2007; Tinwell et al., 2010) and some of their results could have been explained by alternative explanations (e.g., audiovisual asynchrony; Tinwell et al., 2010, in press). We were able to identify only one published study without such outlier characters (Flach et al., 2012) that could be taken as tentative evidence for H1a. This study demonstrated an uncanny curve for experienced discomfort (measured as a dichotomous variable) across video game and film characters that represented different human-likeness levels. We considered this evidence tentative because no statistical tests had been used; furthermore, the human-likeness range was somewhat constrained by the use of only CG characters.

Movement Hypotheses

We were able to identify only two studies (Thompson et al., 2011; Piwek et al., 2014) that could be taken as evidence for the independent movement hypotheses H2a and H2b (**Table 4**). Results from these two studies were, however, consistent with those of the more general hypotheses H1a–b. That is, more humanlike movement was found to elicit higher affinity (H2b) in both studies, whereas a nonlinear uncanny valley curve (H2a) was not observed in either one of them. No studies addressing the modulatory effect of movement (H2c) survived the initial selection and further evaluation. Two studies demonstrated modulatory movement effects; however, these effects were specific to plausible outlier characters (ill-looking face in McDonnell et al., 2012; and zombie character in Piwek et al., 2014). Furthermore, these studies provided conflicting evidence: the former reported a significant increase and the latter a significant decrease in negative affinity for the moving characters.

Categorization Ambiguity Hypotheses

Empirical evidence for categorization ambiguity (H3) and perceptual mismatch (H4) hypotheses is presented in **Table 5**. Four studies demonstrated that a category boundary existed for the identification of morphed facial image continua (H3a) and three of these studies additionally demonstrated that discrimination performance reached its peak when the images straddled this category boundary (H3b). The opposite prediction that discrimination performance would be the poorest in the vicinity of category boundary (H3d) was not supported by any study. These results hence provided reasonable evidence for the categorical perception of morphed human-likeness continua. In contrast, we managed to identify only two studies that tested affinity responses elicited by categorization ambiguity (H3c); neither of which could be taken as evidence in favor of this hypothesis. Opposite to hypothesis H3e, one study (Cheetham et al., 2014) demonstrated that increased perceptual discrimination difficulty is associated with positive rather than negative affinity.

Two other studies demonstrating favorable evidence for H3c were excluded from the total count because of plausible threats to validity. One image morphing study (Yamada et al., 2013) demonstrated that the slowest identification task response times and the most negative likability evaluations coincided with each other; however, these results were excluded because the likability evaluations could plausibly have been influenced by morphing

TABLE 5 | Empirical evidence for hypotheses H3 (categorization ambiguity) and H4 (perceptual mismatch).

Author/year	H3a	H3b	H3c	H3d	H3e	H4a	H4b
Seyama and Nagayama, 2007						+	+
MacDorman et al., 2009						+	+
Looser and Wheatley, 2010	+	+	−	−			
Cheetham et al., 2011	+	+		−			
Mitchell et al., 2011						+	
Gray and Wegner, 2012						+	
Yamada et al., 2013	(+)		(+)				
Burleigh et al., 2013			(+)				−
Cheetham et al., 2013	+						
Cheetham et al., 2014	+	+	−	−	−		
Mäkäräinen et al., 2014							+
Total	4	3	2	3	1	4	4
+	4	3	0	0	0	4	3
−	0	0	2	3	1	0	1

Conclusions: "+": significant in favor of the hypothesis, and "−": non-significant or significant against the hypothesis. Conclusions in parentheses have been omitted from total scores because of plausible threats to validity. Hypotheses: H3a—category identification, H3b—perceptual discrimination, H3c—categorical identification difficulty, H3d—opposite perceptual discrimination, H3e—perceptual discrimination difficulty, H4a—inconsistent human-likeness, H4b—atypicality.

artifacts (cf. Section Naïve and Morbidity Hypotheses). Consistently, two participants in this study had reported spontaneously after the experiment that "they [had] evaluated the likability of the images based on the presence or absence of morphing noise" (ibid., 4). A more systematic evaluation would be necessary for deciding this issue, however. Another study (Study II in Burleigh et al., 2013) demonstrated that intermediate CG modifications between a goat-like and a fully humanlike face elicited the most eerie and unpleasant evaluations. This result was, however, not taken as evidence for the artificial–human categorization ambiguity (H3c) because the presence of categorization boundary was not tested explicitly. The reported positive uncanny valley finding is nevertheless important in the present context, because it could be interpreted as evidence that some human-likeness manipulations can lead to the uncanny valley. This finding was not included as additional evidence for the hypothesis H1a, however, because several other human-likeness manipulations in this study (Burleigh et al., 2013) did not lead to similar findings.

Perceptual Mismatch Hypotheses

As illustrated in **Table 5**, the results provided good support for the perceptual mismatch hypotheses related to both inconsistent realism levels (H4a; 4 out of 4 studies) and sensitivity to atypical features (H4b; 3 out of 4 studies). Two studies (Seyama and Nagayama, 2007; MacDorman et al., 2009) using continuous human-likeness manipulations demonstrated that the most negative affinity evaluations were elicited when the mismatch between the realism of eyes and faces was the greatest (H4a) and when artificially enlarged eyes were paired with the most realistic (fully human) faces (H4b). Two other studies provided further support for H4a. One study (Mitchell et al., 2011), which had used a factorial design between the realism of a face (robot or human) and voice (synthetic or human), demonstrated that mismatched face–voice pairs elicited higher eeriness than similar matched pairs. This result was included as support for H4a, although it should be noticed that these results are somewhat limited because only one pair of stimuli were used in the study. Another study (Gray and Wegner, 2012) with conceptual stimuli demonstrated that machines with characteristically human experiences (i.e., capability to feel) and humans without such experiences were considered unnerving.

Consistently with H4b, one additional study (Mäkäräinen et al., 2014) in **Table 5** demonstrated that unnaturally exaggerated facial expressions were rated as more strange on increasingly humanlike faces. Contrary to H4b, one other study (Burleigh et al., 2013) failed to demonstrate higher eeriness or unpleasantness for increasingly realistic faces. Although this non-significant finding was included in the total count, it is possible that this result could have been specific to the atypical feature (rolled-back eye) used in the study. Unlike enlarged eyes (e.g., Seyama and Nagayama, 2007), for example, such features could appear disturbing both on human and artificial faces.

Some studies that were excluded during the initial selection because they were not fully consistent with the specific formulation of the atypicality hypothesis (H4b) could nevertheless provide further evidence for it. One previous study (Green et al., 2008) demonstrated that individuals show greater agreement when judging the "best looking" facial proportions of human rather than artificial faces. Similar greater agreement for more realistic CG textures was demonstrated also in the second study of MacDorman et al. (2009). Furthermore, the third study in the same article showed that extreme facial proportions were considered the most eerie at close to humanlike levels. These results strengthen the view that individuals are more sensitive and less tolerant to deviations from typical norms when judging human faces.

Discussion

This review considered evidence for the uncanny valley hypothesis (Mori, 1970) based on a framework of specific hypotheses motivated by previous literature. The results showed that whereas all human-likeness manipulations do not automatically lead to the uncanny valley, positive uncanny valley findings have been reported in studies using perceptually mismatching stimuli. In particular, positive uncanny valley findings have been reported for stimuli in which the realism levels of artificial and humanlike features are inconsistent with each other (e.g., human eyes on an artificial face) or in which atypical features (e.g., grossly enlarged eyes) are present on humanlike faces.

Evidence for Different Kinds of Uncanny Valleys

Given that the original uncanny valley formulation did not provide specific guidelines for operationalizing human-likeness, we first considered the straightforward prediction that any kind of successful human-likeness manipulation would lead to the characteristic U-shaped affinity curve at almost humanlike levels. The reviewed studies, which had used various human-likeness manipulations, provided very little support for this hypothesis. Nonlinear uncanny valley effects were found only in two studies that had studied images of hands (Poliakoff et al., 2013 and a continuous CG modification between nonhuman and human faces (Study II in Burleigh et al., 2013). Whether these results could be explained by chance, some characteristics specific to these stimuli or by the other reviewed hypotheses (e.g., categorization ambiguity or perceptual mismatch) remains an open question. The absence of evidence for the naïve uncanny valley hypothesis suggests that all kinds of human-likeness manipulations do not automatically lead to the uncanny valley. This would also suggest that individual studies using only one type of human-likeness manipulation should not be taken as conclusive evidence for the existence or nonexistence of the uncanny valley.

The original uncanny valley formulation also led to the secondary prediction that any kind of human-likeness manipulations would elicit linear increases in experienced affinity. This prediction was supported by the bulk of studies. This suggests that as a general rule, increasing human-likeness is associated with more positive experiences. Exceptions to this general rule could be possible, however, given that different kinds of human-likeness manipulations were not considered systematically in the present review.

We have suggested that Mori used corpses and zombies only as metaphorical examples when discussing threat avoidance as a possible explanation for the uncanny valley. Because these

examples could nevertheless be taken literally, we also considered the hypothesis that such morbid characters would elicit negative affinity. Not surprisingly, this hypothesis received support. The inclusion of this hypothesis was successful because it helped us avoid drawing false conclusions for the other hypotheses. We conclude that empirical studies should not use purposefully morbid characters to test the existence of the uncanny valley (such stimuli could, of course, be included for other purposes). Although another possible confound, purposeful aesthetic, could also have originated from a literal interpretation of the original examples, this issue did not seem to affect any of the reviewed studies.

The original uncanny valley formulation proposed that movement would amplify the characteristic uncanny curve. The reviewed studies did not support this prediction. In contrast, the reviewed studies again demonstrated a linear relationship between affinity and the human-likeness of movement patterns. Furthermore, no nonlinear uncanny valley effects were observed. This suggests that movement information imposes similar linear effects on affinity as any other variation in human-likeness. However, it should be noticed that refined uncanny valley hypotheses (see below) have up to date been studied using only static stimuli, and that movement could possibly amplify their effects.

An alternative claim to the prediction that any kind of human-likeness manipulation leads to the uncanny valley would be that the uncanny valley phenomenon is manifested only under specific conditions. For evaluating this possibility, we considered empirical evidence for two refined uncanny valley proposals as they have been presented in existing literature. First, we considered the claim that the uncanny valley would be caused by an artificial–human categorization ambiguity. Although the reviewed studies demonstrated that morphed artificial–human face continua are perceived categorically, we were able to identify only tentative evidence for negative affinity in the vicinity of category boundary. Taken together, these results suggest that the uncanny valley phenomenon could not be explained solely in terms of categorical perception. However, given the small number of reviewed studies, more conclusive results could yet be obtained in future studies. The uncanny valley hypothesis could also be interpreted such that it predicts greater perceptual discrimination difficulty and more negative affect in the vicinity of category boundary (cf. Cheetham et al., 2014). Neither of these hypotheses was supported by the reviewed evidence.

Second, we considered two different perceptual mismatch hypotheses for the uncanny valley. The first hypothesis predicted that the negative affinity associated with the uncanny valley would be caused by inconsistent realism levels (e.g., artificial eyes on a humanlike face or vice versa). The second hypothesis predicted that such negative affinity would be elicited by heightened sensitivity to atypical features (e.g., grossly enlarged eyes) on humanlike characters. Both of these hypotheses received support from the reviewed studies. This finding is important because it confirms the existence of the uncanny valley at least under some specific conditions. Although previous reviews have presented categorization difficulty and perceptual mismatch hypotheses separately (e.g., Pollick, 2010), we are not aware that a further distinction would have been made between different perceptual mismatch hypotheses. Notably, the reviewed inconsistency and atypicality hypotheses lead to slightly different symmetric and asymmetric predictions. That is, the inconsistency hypothesis would predict that both artificial features on humanlike characters and humanlike features on artificial characters will elicit negative affinity, whereas the atypicality hypothesis would predict atypicality effects only for humanlike stimuli. Because both predictions received support, this suggests that inconsistent realism levels and atypical features could represent different conditions leading to the uncanny valley.

Open Research Questions

The present review raises several open questions for the uncanny valley research. One of these is the relation between the perceptual mismatch and categorization ambiguity hypotheses, which are not necessarily independent from each other. For example, it is possible that realism level inconsistency and feature atypicality effects could be reduced to categorical perception. This idea could possibly be tested by varying the level of inconsistency between features (e.g., by morphing eyes and faces separately as in Seyama and Nagayama, 2007) or by varying the level of feature atypicality (e.g., by varying the eye size of artificial and human faces), and testing whether such continua would fulfill the category identification and perceptual discrimination criteria for categorical perception (Repp, 1984; Harnad, 1987). If these criteria were fulfilled, the results would link these effects to the broader framework of categorical perception.

Another open question relates to whether any kind of perceptual mismatch would lead to the uncanny valley or whether this effect would apply the best or even exclusively to specific features. For example, it might not be a coincidence that two of the reviewed studies demonstrated a perceptual mismatch effect for inconsistent realism levels specifically between the eyes and faces and specifically for enlarged eyes presented on human faces (Seyama and Nagayama, 2007; MacDorman et al., 2009). One of the earliest reviews on the uncanny valley suggested that the eyes would have a special role in producing the uncanny valley (Brenton et al., 2005). Consistently, one image morphing study has demonstrated that human-likeness manipulations of eyes explain most (albeit not all) of the perceived animacy of faces (Looser and Wheatley, 2010). Similarly, one eye tracking study has demonstrated that eyes receive longer gaze dwell time on categorically ambiguous than on categorically unambiguous artificial faces (Cheetham et al., 2013). To our knowledge, the previous suggestion that negative affinity would be caused by inconsistent static and dynamic information (Brenton et al., 2005; Pollick, 2010) also remains unexplored.

The lack of universally agreed upon operational definition for the affinity dimension is a critical issue for uncanny valley studies. The self-report items eeriness, likability, familiarity, and affinity could be derived from Mori's (1970) original formulation. Unfortunately, an inspection of the reviewed articles (Table S1) reveals that none of these single terms alone have been adopted in more than half of the reviewed articles, even after similar terms would be considered as their synonyms (e.g., creepy and strange for eerie; pleasant or appealing for likable; and strange–familiar for familiar). Furthermore, although these items are consistent with

the original formulation, they are not necessarily theoretically justified. One starting point for operationalizing affinity could be the questionnaire developed by Ho and MacDorman (2010). In the present investigation, we have defined affinity in terms of perceptual familiarity and emotional valence. However, these constructs are clearly separate from each other, and their relation in the uncanny valley context would merit further investigation.

Future studies could also consider the possible influences of image morphing artifacts on uncanny valley findings, for example by conducting independent image quality evaluations for morphed stimuli. Although the risk of image morphing artifacts can be diminished considerably by following the guidelines of Cheetham and Jäncke (2013), it is nevertheless possible that all confounding factors would not be avoided. Specifically, some ghosting for subtle facial features that are present in only one of the original images and slight blurring of contours generated by color interpolation could be unavoidable. By the nature of image morphing procedure, middle images in the series of morphed images are the most processed (in a technical sense) and hence they differ the most from natural images that constitute the endpoints of the series. Assuming that morphing artifacts were a realistic concern, the level of visual distortions produced by morphing would hence increase toward the middle of the generated human-likeness continua. The effects of such visual distortions would likely depend on the adopted research question and experimental design, however. Visual distortions, which would likely elicit negative evaluations, could lead to false negative affinity findings at the middle of the scale. On the other hand, it seems unlikely that visual distortions would explain the enhanced discrimination of stimuli straddling the scale middle (i.e., category boundary), as has been reported in typical categorical perception studies. If discrimination were based on comparing visual distortion levels, discrimination should on the contrary be enhanced for adjacent images that are located on either the left or right sides of the scale middle (i.e., for images with different distortion levels) but decreased for images that straddle the scale middle (i.e., for images with symmetric distortion levels).

Limitations

A plausible limitation related to our conceptual analysis of the original uncanny valley formulation (Mori, 1970) is that we have relied on its English translation and other secondary sources instead of the original article written in Japanese.

Given our inclusion criteria, we have only considered studies that have operationalized affinity by self-report measures. We acknowledge that the heterogeneity of self-report items used in the previous studies has significantly reduced the value of comparing their results with one another. Another consequence is that we have omitted several relevant studies that have used physiological and behavioral measures, such as gaze tracking (e.g., Shimada et al., 2006) and haemodynamic response measurements in the brain (e.g., Chaminade et al., 2007; Saygin et al., 2012). It could also be argued that identification task response times, which have already been utilized in some categorical studies (e.g., Looser and Wheatley, 2010; Cheetham et al., 2011), would in fact be good operational definitions of perceptual familiarity. A justification for the present focus on self-report measurements is that their results are easier to interpret than those of physiological or behavioral measures. On the other hand, it should be acknowledged that physiological and behavioral measures could possibly avoid the present ambiguities related to self-report items.

The present conclusions depend on the adopted evaluation criteria, which are to some extent open to subjective interpretations. The function of these criteria was to avoid drawing false conclusions for our hypotheses; consequently, the criteria focused on plausible threats to conclusions that could be drawn from individual studies. We have attempted to facilitate the critical evaluation of this procedure by making it as transparent as possible. Because all possible aspects of experimental validity were not covered, the adopted criteria cannot and should not be taken as evidence for the experimental validity of the evaluated studies themselves. It should also be noticed that although we have specified the minimal human-likeness levels required for testing each hypothesis, this has been done solely for covering as many studies as possible. These minimal levels should hence not be taken as practical guidelines for empirical studies.

Although we have considered only the categorization ambiguity and perceptual mismatch explanations for the uncanny valley, it is worth noting that several other explanations have also been suggested (e.g., see MacDorman and Ishiguro, 2006). For example, it has been suggested that realistic appearance would elicit unrealistic cognitive expectations (expectation violation); that non-lifelike characters would trigger innate fear of death (terror management); and that some artificial characters would be eerie because they appear unfit, infertile, ill, or elicit other evolutionarily motivated aversive responses (evolutionary aesthetics). These explanations operate at different levels—the first two refer to proximate causes (i.e., how the uncanny valley is caused), whereas the evolutionary explanation refers to an ultimate cause (why the uncanny valley exists; cf. Scott-Phillips et al., 2011). Other refinements of the uncanny valley theory have suggested, for example, that behavior that is consistent with a character's appearance will lead to more positive reactions (i.e., a synergy effect; Minato et al., 2004; Ishiguro, 2006). Although these are all empirically testable hypotheses, we have not included them in the present review because they are either similar to the already included hypotheses (e.g., expectation violation vs. inconsistent realism hypotheses) or because they address higher-level topics that seem to presuppose the existence of the uncanny valley in one form or another.

We also acknowledge a recent refinement of the categorization ambiguity hypothesis, which has been suggested in two other articles of the present *Frontiers* Research Topic. As discussed by Schoenherr and Burleigh (2015), the uncanny valley could represent an overarching "inverse mere-exposure effect" (ibid., 3), in which negative affect is caused by a lack of exposure to specific stimuli or stimulus categories (e.g., the authors cite the octopus as a species that is mundanely difficult to categorize). Burleigh and Schoenherr (2015) extend this idea by demonstrating that categorization ambiguity and the frequency of exposure to specific within-category stimuli contribute independently to the uncanny valley. For example, novel stimuli that were extrapolations of their original training stimuli were categorized easily but were nevertheless considered more eerie than stimuli within

their training set. These recent considerations suggest that the categorization ambiguity hypothesis alone would not necessarily be sufficient for predicting emotional responses to the uncanny valley.

Importance and Implications for Research and Practice

Previous articles have already reviewed the uncanny valley phenomenon (e.g., Brenton et al., 2005; Gee et al., 2005; Hanson, 2005; Ishiguro, 2007; Eberle, 2009; Pollick, 2010; Tondu and Bardou, 2011; Zlotowski et al., 2013) and explicated, for example, the categorization ambiguity (e.g., Cheetham et al., 2011) and perceptual mismatch (e.g., MacDorman et al., 2009) hypotheses. However, to our knowledge, this article is the first systematic review of the empirical evidence for the uncanny valley. Conceptual analysis of the uncanny valley and consideration of plausible threats to the conclusions drawn from previous studies to the present hypotheses were used to improve the accuracy of our conclusions. The main contribution of the present article is the conclusion that all kinds of imaginable human-likeness manipulations do not automatically lead to the uncanny valley.

The practical implications of the present findings for computer animators and human-computer or human-robot interaction developers hinge on whether these findings can be generalized to realistic stimuli and contexts—that is, whether they are externally valid (the somewhat redundant term ecological validity could also be used; cf. Kvavilashvili and Ellis, 2004). The present review failed to identify direct evidence for or against the uncanny valley in realistic stimuli, with the exception of some tentative findings (Flach et al., 2012; for other excluded but relevant studies, see Schneider et al., 2007; Tinwell, 2009; Tinwell et al., 2010). However, the reviewed results for artificial but well-controlled stimuli should be generalizable to computer animations and other realistic stimuli as well, given that the experimental stimuli clearly represented phenomena that would be likely to exist also in the real world (cf. Kvavilashvili and Ellis, 2004). For example, it is easy to imagine real computer-animated characters, whose individual features differ from each other with respect to their realism (i.e., perceptual mismatch due to inconsistent realism).

The present results could be taken to encourage the development of increasingly realistic computer animations (and other artificial characters), given that more humanlike characters were in general found to elicit more positive affinity. However, the perceptual mismatch results suggest that the uncanny valley remains a plausible threat for such characters. A generally humanlike character with subtle flaws in some focal features (e.g., eyes), would be likely to elicit negative affinity. The reviewed findings that individuals are increasingly sensitive to atypical features on more humanlike characters would suggest that avoiding the uncanny valley will become exponentially more difficult as the characters' overall appearance approaches the level of full human-likeness. This does not mean that computer animators or robotics researchers should shy away from the grand challenge of creating fully humanlike artificial entities. However, for many practical applications, there may be certain wisdom in the Mori's (1970) original advice of escaping the uncanny valley by attempting to design only moderately humanlike entities.

Conclusion

Taken together, the present review suggested that although not any kind of human-likeness manipulation leads to the uncanny valley, the uncanny valley could be caused by more specific perceptual mismatch conditions. Such conditions could originate, at least, from inconsistent realism levels between individual features (e.g., artificial eyes on a humanlike face) or from the presence of atypical features (e.g., atypically large eyes) on an otherwise humanlike character. Categorical perception of human-likeness continua ranging from artificial to human was supported; however, the present findings failed to support the suggestion that categorization ambiguity would be associated with experienced negative affinity. The results also highlight the need for developing a unified metric for evaluating the subjective, perceptual, and emotional experiences associated with the uncanny valley.

Acknowledgments

This work has been in part supported by the HeCSE and UCIT graduate schools. We thank Dr. Pia Tikka, Prof. Niklas Ravaja, and Dr. Aline de Borst for fruitful discussions on the topic.

Supplementary Material

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpsyg.2015.00390/abstract

References

Alexander, O., Rogers, M., Lambeth, W., Chiang, J.-Y., Ma, W.-C., Wang, C.-C., et al. (2010). The Digital Emily project: achieving a photorealistic digital actor. *IEEE Comput. Graph. Appl.* 30, 20–31. doi: 10.1109/MCG.2010.65

Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. (2007). "Is the uncanny valley an uncanny cliff?" in *Proceedings of the IEEE 16th International Workshop on Robot and Human Interactive Communication (RO-MAN)* (Jeju), 368–373.

Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. (2009). "My robotic doppelgänger—a critical look at the uncanny valley," in *Proceedings of the IEEE 18th International Workshop on Robot and Human Interactive Communication (RO-MAN)* (Toyama), 269–276.

Bradley, M. M., and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25, 49–59. doi: 10.1016/0005-7916(94)90063-9

Brenton, H., Gillies, M., Ballin, D., and Chatting, D. (2005). "The uncanny valley: does it exist?" in *Proceedings of the 19th British HCI Group Annual Conference* (Edinburgh).

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

Burleigh, T. J., Schoenherr, J. R., and 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. *Comput. Hum. Behav.* 29, 759–771. doi: 10.1016/j.chb.2012.11.021

Campbell, R., Pascalis, O., Coleman, M., Wallace, S. B., and Benson, P. J. (1997). Are faces of different species perceived categorically by human observers? *Proc. R. Soc. Lond. B* 264, 1429–1434. doi: 10.1098/rspb.1997.0199

Carter, E. J., Mahler, M., and Hodgins, J. K. (2013). "Unpleasantness of animated characters corresponds to increased viewer attention to faces," in *Proceedings of the ACM Symposium on Applied Perception (SAP)* (Dublin), 35–40.

Chaminade, T., Hodgins, J., and Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters' actions. *Soc. Cogn. Affect. Neurosci.* 2, 206–216. doi: 10.1093/scan/nsm017

Cheetham, M., and Jäncke, L. (2013). Perceptual and category processing of the Uncanny Valley hypothesis' dimension of human-likeness: some methodological issues. *J. Vis. Exp.* 76:e4375. doi: 10.3791/4375

Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., and Jäncke, L. (2013). Category processing and the human likeness dimension of the uncanny valley hypothesis: eye-tracking data. *Front. Psychol.* 4:108. doi: 10.3389/fpsyg.2013.00108

Cheetham, M., Suter, P., and Jäncke, L. (2011). The human likeness dimension of the "uncanny valley hypothesis": behavioral and functional MRI findings. *Front. Hum. Neurosci.* 5:126. doi: 10.3389/fnhum.2011.00126

Cheetham, M., Suter, P., and Jäncke, L. (2014). Perceptual discrimination difficulty and familiarity in the Uncanny Valley: more like a "Happy Valley." *Front. Psychol.* 5:1219. doi: 10.3389/fpsyg.2014.01219

de Gelder, B., Teunisse, J.-P., and Benson, P. J. (1997). Categorical perception of facial expressions: categories and their internal structure. *Cogn. Emot.* 11, 1–23. doi: 10.1080/026999397380005

Eberle, S. G. (2009). Exploring the uncanny valley to find the edge of play. *Am. J. Play* 2, 167–194.

Flach, L. M., de Moura, R. H., Musse, S. R., Dill, V., Pinho, M. S., and Lykawka, C. (2012). "Evaluation of the uncanny valley in CG characters," in *Proceedings of the Brazilian Symposium on Computer Games and Digital Entertainmen (SBGames)* (Brasiilia), 108–116.

Gee, F. C., Browne, W. N., and Kawamura, K. (2005). "Uncanny valley revisited," in *Proceedings of the 14th IEEE International Workshop on Robot and Human Interactive Communication* (Nashville, TN), 151–157.

Geller, T. (2008). Overcoming the uncanny valley. *IEEE Comput. Graph. Appl.* 28, 11–17. doi: 10.1109/MCG.2008.79

Goldstone, R. L., and Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 69–78, doi: 10.1002/wcs.26

Gray, K., and Wegner, D. M. (2012). Feeling robots and human zombies: mind perception and the uncanny valley. *Cognition* 125, 125–130. doi: 10.1016/j.cognition.2012.06.007

Green, B. F., and Hall, J. A. (1984). Quantitative methods for literature reviews. *Ann. Rev. Psychol.* 35, 35–53. doi: 10.1146/annurev.ps.35.020184.000345

Green, R. D., MacDorman, K. F., Ho, C.-C., and Vasudevan, S. (2008). Sensitivity to the proportions of faces that vary in human likeness. *Comput. Hum. Behav.* 24, 2456–2474. doi: 10.1016/j.chb.2008.02.019

Hanson, D. (2005). "Expanding the aesthetic possibilities for humanoid robots," in *Proceedings of the 5th IEEE-RAS International Conference on Humanoid Robots* (Tsukuba).

Hanson, D. (2006). "Exploring the aesthetic range for humanoid robots," in *Proceedings of the 28th Annual Conference of the Cognitive Science Society (CogSci)* (Vancouver, BC), 16–20.

Harnad, S. R. (1987). "Introduction: psychophysical and cognitive aspects of categorical perception: a critical overview," in *Categorical Perception: The Groundwork of Cognition,* ed S. R. Harnad (Cambridge: Cambridge University Press), 1–29.

Ho, C.-C., and MacDorman, K. F. (2010). Revisiting the uncanny valley theory: developing and validating an alternative to the godspeed indices. *Comput. Hum. Behav.* 26, 1508–1518. doi: 10.1016/j.chb.2010.05.015

Ho, C.-C., MacDorman, K. F., and Pramono, Z. A. W. (2008). "Human emotion and the uncanny valley: a GLM, MDS, and isomap analysis of robot video ratings," in *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (Amsterdam), 169–176.

Ishiguro, H. (2006). Android science—toward a new cross-interdisciplinary framework. *Interact. Stud.* 7, 297–337. doi: 10.1075/is.7.03mac

Ishiguro, H. (2007). Scientific issues concerning androids. *Int. J. Rob. Res.* 26, 105–117. doi: 10.1177/0278364907074474

Kvavilashvili, L., and Ellis, J. (2004). Ecological validity and the real-life/laboratory controversy in memory research: a critical (and historical) review. *Hist. Phil. Psychol.* 6, 59–80.

Langlois, J. H., and Roggman, L. A. (1990). Attractive faces are only average. *Psychol. Sci.* 1, 115–121. doi: 10.1111/j.1467-9280.1990.tb00079.x

Looser, C. E., and Wheatley, T. (2010). The tipping point of animacy: how, when, and where we perceive life in a face. *Psychol. Sci.* 21, 1854–1862. doi: 10.1177/0956797610388044

MacDorman, K. F. (2005). "Androids as experimental apparatus: why is there an uncanny valley and can we exploit it?" in *Proceedings of the Cognitive Science Society (CogSci) Workshop on Toward Social Mechanisms of Android Science* (Stresa), 108–118.

MacDorman, K. F. (2006). "Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: an exploration of the uncanny valley," in *Proceedings of the 28th Annual Conference of the Cognitive Science Society (CogSci)* (Vancouver), 26–29.

MacDorman, K. F., Green, R. D., Ho, C.-C., and Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Comput. Hum. Behav.* 25, 695–710. doi: 10.1016/j.chb.2008.12.026

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

MacDorman, K. F., Srinivas, P., and Patel, H. (2013). The uncanny valley does not interfere with level-1 visual perspective taking. *Comput. Hum. Behav.* 29, 1671–1685. doi: 10.1016/j.chb.2013.01.051

Mäkäräinen, M., Kätsyri, J., and Takala, T. (2014). Exaggerating facial expressions: a way to intensify emotion or a way to the uncanny valley? *Cogn. Comput.* 6, 708–721. doi: 10.1007/s12559-014-9273-0

Matsuda, Y.-T., Okamoto, Y., Ida, M., Okanoya, K., and Myowa-Yamakoshi, M. (2012). Infants prefer the faces of strangers or mothers to morphed faces: an uncanny valley between social novelty and familiarity. *Biol. Lett.* 8, 725–728. doi: 10.1098/rsbl.2012.0346

McDonnell, R., Breidt, M., and Bülthoff, H. H. (2012). Render me real? Investigating the effect of render style on the perception of animated virtual humans. *ACM Trans. Graph.* 31, 1–11. doi: 10.1145/2185520.2185587

Minato, T., Shimada, M., Ishiguro, H., and Itakura, S. (2004). Development of an android robot for studying human-robot interaction. *Lect. Notes Comput. Sci.* 3029, 424–434. doi: 10.1007/978-3-540-24677-0_44

Misselhorn, C. (2009). Empathy with inanimate objects and the uncanny valley. *Mind Mach.* 19, 345–359. doi: 10.1007/s11023-009-9158-2

Mitchell, W. J., Szerszen, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., and MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *I-Perception* 2, 10–12. doi: 10.1068/i0415

Moore, R. K. (2012). A bayesian explanation of the "Uncanny Valley" effect and related psychological phenomena. *Sci. Rep.* 2:864. doi: 10.1038/srep00864

Mori, M. (1970). Bukimi no tani [the uncanny valley]. *Energy* 7, 33–35. Transl. K. F. MacDorman and N. Kageki (2012), *IEEE Trans. Rob. Autom.* 19, 98–100. doi: 10.1109/MRA.2012.2192811

Perry, T. S. (2014). Leaving the uncanny valley behind. *IEEE Spectr.* 51, 48–53. doi: 10.1109/MSPEC.2014.6821621

Pisoni, D. B., and Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Percept. Psychophys.* 15, 285–290. doi: 10.3758/BF03213946

Piwek, L., McKay, L. S., and Pollick, F. E. (2014). Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition* 130, 271–277. doi: 10.1016/j.cognition.2013.11.001

Poliakoff, E., Beach, N., Best, R., Howard, T., and Gowen, E. (2013). Can looking at a hand make your skin crawl? Peering into the uncanny valley for hands. *Perception* 42, 998–1000. doi: 10.1068/p7569

Pollick, F. E. (2010). In search of the uncanny valley. *Lect. Note Inst. Comput. Sci. Telecomm.* 40, 69–78. doi: 10.1007/978-3-642-12630-7_8

Ramey, C. H. (2005). "The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robot," in *Proceedings of the 5th IEEE-RAS International Conference on Humanoid Robots* (Tsukuba).

Repp, B. H. (1984). "Categorical perception: issues, methods, findings," in *Speech and Language: Advances in Basic Research and Practise* Vol. 10, ed. N. J. Lass (Orlando: Academic Press), 243–335.

Rhodes, G., Yoshikawa, S., Clark, A., Lee, K., McKay, R., and Akamatsu, S. (2001). Attractiveness of facial averageness and symmetry in non-Western cultures: in search of biologically based standards of beauty. *Perception* 30, 611–625. doi: 10.1068/p3123

Rosenthal–von der Pütten, R., and Krämer, N. C. (2014). How design characteristics of robots determine evaluation and uncanny valley related responses. *Comput. Hum. Behav.* 36, 422–439. doi: 10.1016/j.chb.2014.03.066

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychol. Rev.* 110, 145–172. doi: 10.1037/0033-295X.110.1.145

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., and Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* 7, 413–422. doi: 10.1093/scan/nsr025

Schneider, E., Wang, Y., and Yang, S. (2007). "Exploring the uncanny valley with Japanese video game characters," in *Proceedings of the Digital Games Research Association (DiGRA): Situated Play*, ed B. Akira (Tokyo), 546–549.

Schoenherr, J. R., and Burleigh, T. J. (2015). Uncanny sociocultural categories. *Front. Psychol.* 5:1456. doi: 10.3389/fpsyg.2014.01456

Scott-Phillips, T. C., Dickins, T. E., and West, S. A. (2011). Evolutionary theory and the ultimate-proximate distinction in the human behavioral sciences. *Perspect. Psychol. Sci.* 6, 38–47. doi: 10.1177/1745691610393528

Seyama, J., and Nagayama, R. S. (2007). The uncanny valley: effect of realism on the impression of artificial human faces. *Presence* 16, 337–351. doi: 10.1162/pres.16.4.337

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference.* Belmont, CA: Wadsworth.

Shimada, M., Minato, T., Itakura, S., and Ishiguro, H. (2006). "Evaluation of android using unconscious recognition," in *Proceedings of the IEEE-RAS International Conference on Humanoid Robots* (Genova), 157–162.

Steckenfinger, S. A., and Ghazanfar, A. A. (2009). Monkey visual behavior falls into the uncanny valley. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18362–18366. doi: 10.1073/pnas.0910063106

Thompson, J. C., Trafton, J. G., and McKnight, P. (2011). The perception of humanness from the movements of synthetic agents. *Perception* 40, 695–704. doi: 10.1068/p6900

Tinwell, A. (2009). Uncanny as usability obstacle. *Lect. Notes Comput. Sci.* 5621, 622–631. doi: 10.1007/978-3-642-02774-1_67

Tinwell, A., Grimshaw, M., and Nabi, D. A. (in press). The effect of onset asynchrony in audio-visual speech and the uncanny valley in virtual characters. *Int. J. Digital Hum.* 2.

Tinwell, A., Grimshaw, M., and Williams, A. (2010). Uncanny behaviour in survival horror games. *J. Gaming Virtual Worlds* 2, 3–25. doi: 10.1386/jgvw.2.1.3_1

Tondu, B., and Bardou, N. (2011). A new interpretation of Mori's uncanny valley for future humanoid robots. *Int. J. Robot. Autom.* 26, 337–348. doi: 10.2316/Journal.206.2011.3.206-3348

Wolberg, G. (1998). Image morphing: a survey. *Vis. Comput.* 14, 360–372. doi: 10.1007/s003710050148

Wu, E., and Liu, F. (2013). Robust image metamorphosis immune from ghost and blur. *Vis. Comput.* 29, 311–321. doi: 10.1007/s00371-012-0734-8

Yamada, Y., Kawabe, T., and Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *Jpn. Psychol. Res.* 55, 20–32. doi: 10.1111/j.1468-5884.2012.00538.x

Zlotowski, J., Proudfoot, D., Bartneck, C., and Złotowski, J. (2013). "More human than human: does the uncanny curve really matter?" in *Proceedings of the HRI2013 Workshop on Design of Humanlikeness in HRI* (Tokyo), 7–13.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 13 March 2015
doi: 10.3389/fpsyg.2015.00249

Stimulus-category competition, inhibition, and affective devaluation: a novel account of the uncanny valley

Anne E. Ferrey[1], *Tyler J. Burleigh*[2] *and Mark J. Fenske*[2]*

[1] *Child Study Center, Yale University School of Medicine, New Haven, CT, USA,* [2] *Department of Psychology, University of Guelph, Guelph, ON, Canada*

Stimuli that resemble humans, but are not perfectly human-like, are disliked compared to distinctly human and non-human stimuli. Accounts of this "Uncanny Valley" effect often focus on how changes in human resemblance can evoke different emotional responses. We present an alternate account based on the novel hypothesis that the Uncanny Valley is not directly related to 'human-likeness' *per se,* but instead reflects a more general form of stimulus devaluation that occurs when inhibition is triggered to resolve conflict between competing stimulus-related representations. We consider existing support for this inhibitory-devaluation hypothesis and further assess its feasibility through tests of two corresponding predictions that arise from the link between conflict-resolving inhibition and aversive response: (1) that the pronounced disliking of Uncanny-type stimuli will occur for any image that strongly activates multiple competing stimulus representations, even in the absence of any human-likeness, and (2) that the negative peak of an 'Uncanny Valley' should occur at the point of greatest stimulus-related conflict and not (in the presence of human-likeness) always closer to the 'human' end of a perceptual continuum. We measured affective responses to a set of line drawings representing non-human animal–animal morphs, in which each continuum midpoint was a bistable image (Experiment 1), as well as to sets of human-robot and human-animal computer-generated morphs (Experiment 2). Affective trends depicting classic Uncanny Valley functions occurred for all continua, including the non-human stimuli. Images at continua midpoints elicited significantly more negative affect than images at endpoints, even when the continua included a human endpoint. This illustrates the feasibility of the inhibitory-devaluation hypothesis and the need for further research into the possibility that the strong dislike of Uncanny-type stimuli reflects the negative affective consequences of cognitive inhibition.

Keywords: uncanny valley, cognitive conflict, inhibition, affect, emotion, inhibitory devaluation, visual perception, cognitive dissonance

OPEN ACCESS

Edited by:
Marcus Cheetham, University of Zürich, Switzerland

Reviewed by:
Rick Thomas, University of Oklahoma, USA
Ute Schmid, University of Bamberg, Germany

***Correspondence:**
Mark J. Fenske, Department of Psychology, University of Guelph, Guelph, ON N1G 2W1, Canada
mfenske@uoguelph.ca

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *30 March 2014*
Accepted: *18 February 2015*
Published: *13 March 2015*

Citation:
Ferrey AE, Burleigh TJ and Fenske MJ (2015) Stimulus-category competition, inhibition, and affective devaluation: a novel account of the uncanny valley. Front. Psychol. 6:249. doi: 10.3389/fpsyg.2015.00249

Introduction

The Uncanny Valley— a significant decrease in liking for objects that closely resemble humans but are not perfectly human-like— was originally described in terms of the uncomfortable feeling associated with viewing robots of increasing human-likeness

(Mori, 1970). Indeed, many accounts of this effect have focused on the potential relationship between the subjective human-likeness of a stimulus and an observer's emotional response to it (e.g., MacDorman and Ishiguro, 2006; Seyama and Nagayama, 2007; MacDorman et al., 2009). The importance of elucidating specific mechanisms underlying the Uncanny Valley effect is underscored by the extent to which interest in this effect has spread from robotics into other areas, such as computer graphics and prosthetics (Seyama and Nagayama, 2007; MacDorman et al., 2009; Mitchell et al., 2011; Tinwell et al., 2011; Poliakoff et al., 2013). To this end, we consider here the novel hypothesis that the Uncanny Valley is not directly related to 'human-likeness' *per se,* but instead reflects a more general form of stimulus devaluation that occurs when inhibition is triggered to resolve conflict between competing stimulus-related representations. The purpose of this article is to therefore demonstrate how the Uncanny Valley may be explained through recent advances in our understanding of the negative affective consequences of cognitive inhibition. After presenting our 'inhibitory-devaluation' hypothesis, we report a preliminary assessment of its feasibility—in terms of prior findings as well as through two new experiments that test specific predictions arising from this new account of the Uncanny Valley—and then consider directions for future research.

Inhibition, Negative Affect, and the Uncanny Valley

A recent and growing body of research suggests that cognitive inhibition is not only crucial for resolving potential interference during visual tasks (i.e., when multiple stimulus/response representations compete to become the focus of thoughts and actions), but also subsequently results in negative affect for the associated stimuli (for reviews, see Fenske and Raymond, 2006; Raymond, 2009). Such affectively negative consequences of inhibition have been found in a variety of visual-recognition tasks that require stimulus classification (e.g., Kiss et al., 2008; Frischen et al., 2012) and localization (e.g., Raymond et al., 2003; Fenske et al., 2004), using stimuli ranging from meaningless patterns (e.g., Raymond et al., 2003), non-human objects (e.g., Griffiths and Mitchell, 2008), and entire scenes (Frischen et al., 2012), to images of real human faces (Fenske et al., 2005), and bodies (Ferrey et al., 2012). Moreover, these studies have shown that this inhibitory devaluation impacts a variety of subjective emotional judgments (i.e., likeability, relative preference, cheerfulness, pleasantness, trustworthiness, sexual attractiveness), as well as the motivational incentive to seek and obtain otherwise-appealing stimuli. Importantly, the magnitude of inhibitory devaluation increases with the level of potential interference from competing stimulus-category or stimulus-response representations (e.g., Raymond et al., 2005; Frischen et al., 2012; Martiny-Huenger et al., 2013). This suggests that the Uncanny Valley effect could be a specific instance of inhibition-related devaluation of stimuli whose perception activates multiple, competing stimulus interpretations.

Most prior accounts of the Uncanny Valley effect suggest it occurs when humans view images of conspecifics (i.e., other potential humans) that possess non-human traits. For instance, one idea is that disliking of not-quite-humanlike images is the result of a disgust response that evolved for the purpose of pathogen avoidance (MacDorman and Ishiguro, 2006; MacDorman et al., 2009). From this perspective, the stronger an entity's resemblance to a conspecific, the stronger the aversion to a "deformed" version would be, since defects may cue potential disease, and conspecific resemblance cues the potential for catching the disease due to genetic similarity (Rozin and Fallon, 1987). By extension, negative feelings elicited during studies that have used human-like stimuli with mismatched features (e.g., Seyama and Nagayama, 2007; MacDorman et al., 2009; Mitchell et al., 2011) could reflect the activation of this human-specific pathogen-avoidance mechanism.

In contrast to theories focusing on discrepancies related to the 'human-ness' of stimuli, an inhibitory-devaluation account of the Uncanny Valley predicts that negative evaluations will be triggered by any stimulus that activates multiple, competing stimulus representations during recognition. Recent reviews of the neurocognitive mechanisms underlying the perception of ambiguous sensory information suggest that resolving such competition, and the accompanying perceptual ambiguity, may be achieved through mechanisms that suppress neural activity associated with perceptual features that conflict with the perceptual outcome (e.g., Sterzer et al., 2009). This perspective is consistent with well-accepted biased-competition models of visual processing (e.g., Desimone and Duncan, 1995), whereby each possible interpretation of an ambiguous figure competes for representation across a hierarchical network of visual areas. Top–down signals, such as those associated with selective attention, can bias this neural competition in favor of one perceptual interpretation over another (Meng and Tong, 2004). Evidence that inhibition of competing representations may be one of the mechanisms through which the competition is biased to resolve such conflict (e.g., Munakata et al., 2011) suggests that the well-established negative affective consequences of inhibition should be evident for any stimulus that activates multiple, competing stimulus representations during recognition, just as it is for stimuli associated with other forms of inhibition (Fenske and Raymond, 2006; Raymond, 2009).

Support for a shift in focus from human-specific to more general recognition-related mechanisms can also be seen in recent 'categorization' accounts of the Uncanny Valley (Cheetham et al., 2011; Moore, 2012; Burleigh et al., 2013; Cheetham et al., 2013; Burleigh and Schoenherr, 2015). These accounts generally consider affective response to be a function of stimulus distance from a category boundary (Cheetham et al., 2011, 2014; but see Burleigh and Schoenherr, 2015). A stimulus is easy to classify as 'human' or 'non-human' when it is far from the category boundary along a 'human'/'non-human' continuum. But a stimulus that is at or near the category boundary is difficult to classify because its identity is ambiguous. Cheetham et al. (2011, 2014) provided evidence of this category boundary at the midpoint of a human-avatar morph continuum, and Burleigh et al. (2013) reported that ambiguous morph-stimuli near the midpoint of a human–non-human continuum were indeed associated with heightened levels

of negative affect. The findings of Cheetham et al. (2014) further support the idea that such devaluation may be linked to categorization, but is not likely to follow from perceptual discrimination difficulty, *per se*. Importantly, these previous results are consistent with the possibility that stimuli near such midpoints strongly activate multiple, competing visual-category representations during recognition. From this perspective, negative affect for such items occurs to the extent that selecting one interpretation over the other requires inhibition of the visual-category information associated with the non-selected interpretation. The greater the inhibition during identification, the greater the negative affect for the associated stimulus.

A key feature of Uncanny Valley explanations that focus on the 'human-ness' of stimuli concerns the expected location of the 'valley' – the point along a perceptual continuum where stimulus-related affective response maximally deviates from an otherwise linear function. A conspecific pathogen-avoidance account, for example, predicts an asymmetrical valley that drops closer to the 'human' side of a 'human'/'nonhuman' continuum. Indeed, this is exactly what was depicted in Mori's (1970) well-known original figure illustrating the Uncanny Valley. In contrast, an inhibitory-devaluation account predicts the greatest affective drop at the point where multiple competing visual-category representations are most strongly activated. And while this should be at the midpoint for continua anchored by two equally distinct stimulus categories, the exact location for any given continuum will vary depending on parameters such as baseline affective response to the specific endpoints and the perceptual salience of the visual cues denoting each category. This may explain why Seyama and Nagayama (2007, Experiments 3), for example, were able to obtain a valley location comparable to that depicted by Mori (1970), but only after dramatically increasing the size (and thus the perceptual salience) of the discrepant features in otherwise highly human-like stimuli. Other studies utilizing 'human'/'nonhuman' continua have found valleys at the midpoint, and occasionally closer to the 'non-human' endpoint (MacDorman and Ishiguro, 2006; Burleigh et al., 2013). Such findings are consistent with the possibility that the lowest point in the valley is not determined by the human-ness of the stimuli *per se*, but instead occurs at whatever point requires the greatest inhibition of competing visual-category information to select one stimulus interpretation over another.

Inhibition has been proposed as a critical mechanism for resolving conflict and potential interference from competing signals in a variety of cognitive and neural operations (for review, see Munakata et al., 2011). And while evidence of the link between conflict-resolving inhibition and negative affect has only recently begun to accumulate, several lines of research, including Burleigh et al.'s (2013) examination of the Uncanny Valley, have demonstrated the link between situations involving cognitive conflict and negative affect. A classic example is cognitive dissonance theory, which originally described how a negative emotional response is elicited when a person's attitude is at odds with their behavior (Festinger, 1957). In more general terms, cognitive dissonance describes a conflict between two incompatible cognitions (van Veen et al., 2009), which leads to both autonomic arousal and negative affect (Croyle and Cooper, 1983; Losch and Cacioppo, 1990; Elliot and Devine, 1994) and has been described as a state of discomfort and unease (Elliot and Devine, 1994). In order to resolve this conflict, cognitive resources must be devoted to the problem. The corresponding negative affect may be linked in part to the extent that the resources recruited in the face of such conflict include inhibition aimed at reducing the salience of incompatible representations.

A special case of cognitive dissonance is *post-decisional dissonance*. This phenomenon was first observed in Brehm (1956), who noticed that participants who were asked to choose between two similarly valued items had rated selected items more positively than their initial ratings of the same item, and rated the rejected item more negatively. In this case, a conflict between two choices lead to negative affect associated with the unchosen item. This may result from recruitment of cortical areas that inhibit representations of the unchosen option, leading to negative affect (Harmon-Jones, 2004).

Other types of interference or conflict are also associated with negative affect. For example, the interference that is caused by response competition during the Stroop (1935) task when the name of a color is presented in a color that is inconsistent with its identity (e.g., the word "blue" in red ink). When conflict occurs, the dorsal anterior cingulate cortex (dACC; Botvinick et al., 2001; van Veen et al., 2009; Izuma et al., 2013) and insula regions (van Veen et al., 2009) are activated, and individuals experience arousal and feelings of discomfort (Elliot and Devine, 1994; van Veen et al., 2009), which motivates them to engage in a dissonance-reduction strategy. In the case of a Stroop (1935) task, for example, dissonance-reduction is accomplished by biasing inputs such that word names or word colors dominate response selection (Botvinick et al., 2001).

The consistency of prior findings with what we have recently learned about the affective consequences of inhibition suggests that an inhibitory-devaluation account of the Uncanny Valley effect merits further consideration. One way to further assess its feasibility is to begin experimentally testing specific predictions that arise from our account. We report two such tests below as a preliminary example of Uncanny Valley research into the inhibitory-devaluation hypothesis.

Assessing the Feasibility of the Inhibitory-Devaluation Hypothesis

The hypothesis that the Uncanny Valley reflects inhibitory devaluation of stimuli that activate multiple, competing stimulus interpretations during recognition generates a number of testable predictions. The two considered here concern the type of stimuli that can show Uncanny Valley effects and the type of stimuli that are likely to receive the most negative affective evaluations. Our experimental approach for assessing the potential involvement of cognitive inhibition in the Uncanny Valley is the same as that used extensively in studies of inhibitory devaluation. In these prior studies, any type of stimulus—human or non-human— appearing in experimental conditions suspected to

involve cognitive inhibition have consistently received more negative evaluations than stimuli appearing in conditions thought to be relatively free of inhibition (Fenske and Raymond, 2006; Raymond, 2009). Such effects are routinely obtained across substantially different visual classification and response-decision tasks. Indeed, the link between inhibition and stimulus devaluation is sufficiently strong that researchers have begun to take the occurrence of increasingly negative subjective stimulus evaluations as a key indicator of the potential involvement of inhibition at key points within a given task (e.g., Kihara et al., 2011). Thus, while we do not directly measure inhibition *per se*, we instead assess whether differences in affective evaluations for items at different points along a given perceptual continuum are consistent with the expected extent to which inhibition may be applied during the perception of such items. The experiments reported in this section therefore represent an important first step in assessing the feasibility of an inhibitory-devaluation account by (1) confirming that the Uncanny Valley also occurs when humans view distinctly non-human stimuli, and (2) demonstrating that for human-like stimuli, the lowest point of the 'valley' does not always occur on the 'human' side of the perceptual continuum.

The key emphasis on the human-ness of objects in prior accounts of the Uncanny Valley effect may explain why so little work has been done to explore whether Uncanny Valley-type effects also occur with non-human stimuli. To the best of our knowledge, the only published research of this sort currently includes two papers by Yamada et al. (2012, 2013). However, their findings do suggest that the Uncanny Valley can occur for non-human stimuli. Yamada et al. (2012) obtained participants' categorization responses and affective ratings of images that morphed between a tomato and a strawberry. The lowest likability scores for these images coincided with the point of greatest ambiguity in stimulus categorization. Yamada et al. (2013) likewise found that participants' most affectively negative ratings were provided for the most ambiguous stimuli among sets of images that morphed between a real dog and a cartoon dog, a real dog and a stuffed-toy dog, and a cartoon dog and a stuffed-toy dog. Unfortunately, the veracity of these findings remains unclear because of issues associated with the use of relatively small sample sizes (e.g., Yamada et al., 2013, utilized 12 or fewer participants in each experiment), and the possibility that their stimuli included visual artifacts produced by the morphing process that may have had a confounding influence on participants' affective ratings.

Our studies therefore expand upon these important prior findings to provide a converging test of the prediction that Uncanny Valley effects should not be limited to the perception of humans, but should also occur with non-human stimuli. Thus, in Experiment 1, we use bistable line drawings of animals, a type of non-human stimuli that has been specifically designed to activate multiple, competing stimulus interpretations (Fisher, 1967). The sets of line-drawn images we used were specifically chosen for their step-wise differences along a given two-category continuum and because of the availability of normative data regarding the corresponding level of perceptual ambiguity of each item (Verstijnen and Wagemans, 2004).

Accordingly, the conditions in Experiment 1 in which we suspect the greatest involvement of inhibition during perception concern those items whose perceptual features are consistent with multiple conflicting interpretations, such as those near the midpoint of a given continuum. In contrast, recognition of items closer to a continuum endpoint, whose perceptual features are clearly consistent with a single interpretation, should be relatively free of inhibition. Following the experimental approach of prior inhibitory devaluation studies, we therefore expect that any stimulus occurring near the midpoint of such a two-category continuum should receive more negative evaluations than items near the continuum endpoints. Participants in our studies were asked to evaluate stimuli based on their initial emotional reaction to each stimulus using a numerical rating scale. This allowed us to obtain an accurate measure of subjective emotional reactions to non-human stimuli that vary in the extent to which they activate multiple, competing stimulus interpretations.

In addition to responses to non-human stimuli, we also predicted that when assessing affective responses to human-like stimuli, the location of the 'valley'— the point along a perceptual continuum where stimulus-related affective response maximally deviates from an otherwise linear function — should occur near the midpoint of a two-category continuum. We tested this prediction in Experiment 2 by examining differences in individuals' affective responses to sets of 3D computer-modeled images that represent different points along human-to-robot and human-to-animal morphed continua.

Experiment 1: Non-Human Bistable Images

Materials and Methods

Stimuli in this experiment consisted of three different sets of line drawings, each comprising a step-wise continuum of differences in perceptual similarity to two distinct animals. The stimulus at the midpoint of each continuum is a bistable image that can be interpreted as either of the two animals (see **Figure 1** for example). Bistable images have long been of interest (e.g., Fisher, 1967) as stimuli that can support two incompatible interpretations, although the stimulus itself does not change. Such stimuli are specifically created to ensure maximal category conflict for items near the bistable midpoint. Normative stimulus-classification data provided by Verstijnen and Wagemans (2004) previously established that the point of maximal perceptual ambiguity for these stimulus sets was for the item within one step

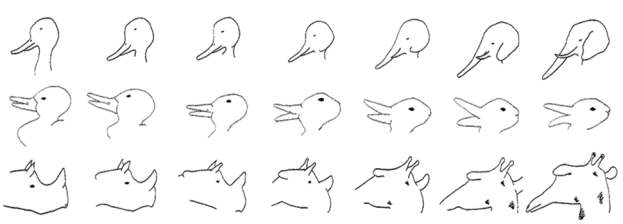


FIGURE 1 | Bistable image examples (rabbit-duck, rhino-giraffe, and duck-elephant; originally from Hogeboom (1995), as cited in Verstijnen and Wagemans (2004).

of the midpoint (i.e., midpoint plus or minus one step) of each continuum.

We predicted that the shape of the affective data for stimuli from each continuum would be consistent with an Uncanny Valley function (i.e., non-linear). Following Burleigh et al. (2013), we tested this by fitting linear, cubic, and quadratic functions to the data for each continuum. Because the Uncanny Valley function (Mori, 1970) is essentially a cubic function, we expected a cubic or quadratic function would fit our data better than a linear function. We also predicted that the specific shape of this non-linear function would be formed by lower affective ratings for morph-stimuli near the midpoint of the continua than for those near the endpoints.

All of the following materials and procedures were approved by the Research Ethics Board at the University of Guelph (REB #11NV011).

Participants

Sixty undergraduate students (31 women, $M_{age} = 20$ years, $SD_{age} = 3.6$) participated in exchange for course credit. The only inclusion criterion was having normal or corrected-to-normal vision.

Apparatus and stimuli

Stimuli consisted of three sets line drawings, each comprising seven images (Hogeboom, 1995, as cited in Verstijnen and Wagemans, 2004). Sets included: duck-elephant, rhino-giraffe, and rabbit-duck (see **Figure 1**).

The 21 images were presented in a randomized order for each participant using an Intel Core2Duo computer with a 50.8 cm LCD monitor (resolution: 1680 × 1050 pixels) running PsychoPy software (Peirce, 2007). Displays were viewed at a distance of 75 cm in a sound-attenuated room, with low ambient illumination. Stimuli were presented one at a time at the center of the screen for 400 ms along with a visual-analog rating scale that ranged from "dislike very much" (0.00) to "like very much" (1.00). Participants were required to use the mouse to select the point on the line that best matched their emotional response to each stimulus. The rating scale had a precision of 0.01 unit increments, and was visible on the screen until a response was registered.

Results

Average ratings were calculated for each stimulus as a function of its location along its corresponding continuum. These are plotted separately for each stimulus set in **Figure 2**, along with average ratings calculated across the three stimulus sets.

We predicted that the shape of the affective data for stimuli from each continuum would be consistent with an Uncanny Valley function (i.e., non-linear). To assess this, we fit linear, cubic, and quadratic functions to the data for each continuum, as in Burleigh et al. (2013). Because the Uncanny Valley function (Mori, 1970) is essentially a cubic function, it follows that

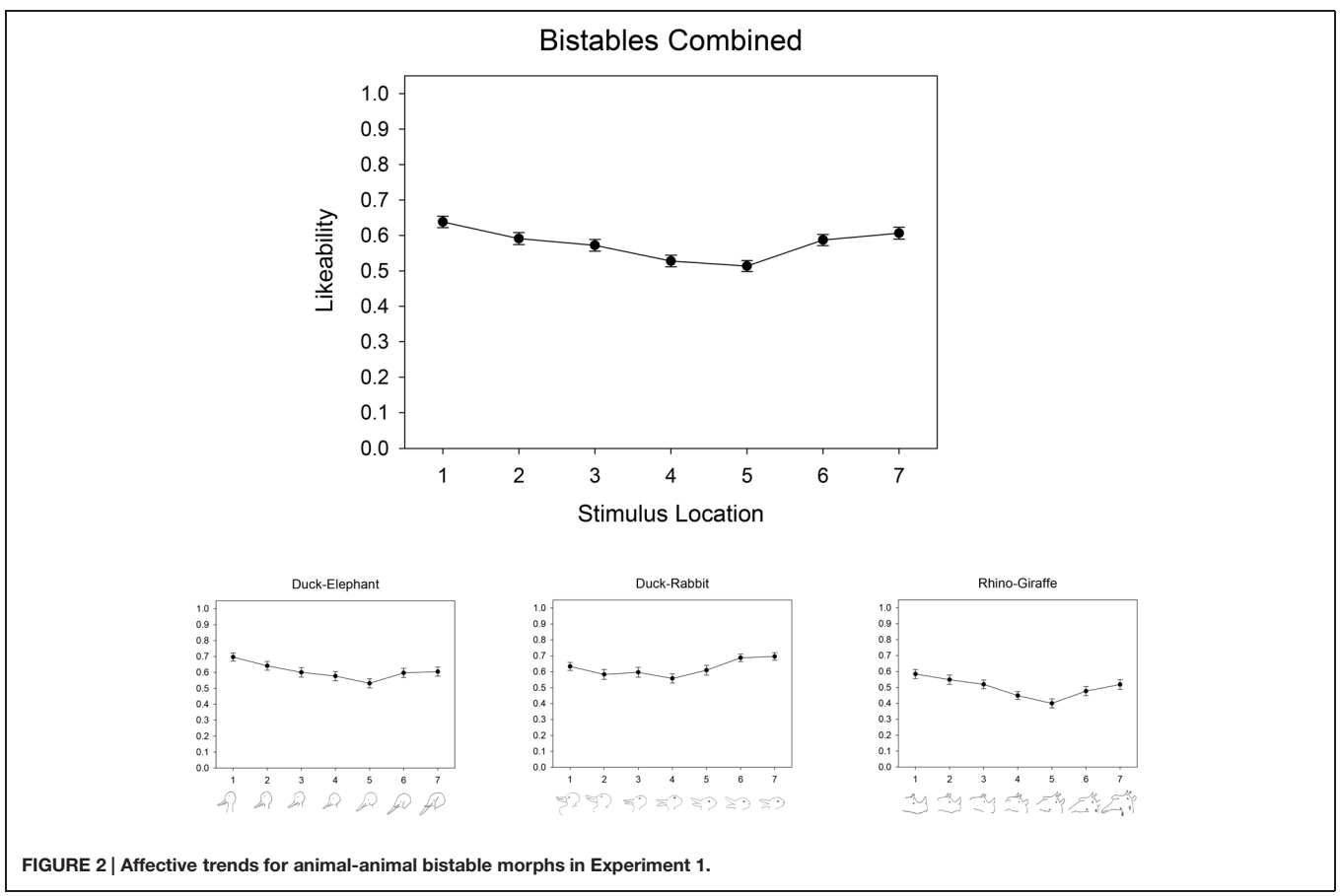


FIGURE 2 | Affective trends for animal-animal bistable morphs in Experiment 1.

if a cubic or quadratic function was found to fit the data better than a linear function, then this would support an Uncanny Valley interpretation.

We used the Akaike Information Criterion (AIC; see Burnham and Anderson, 2002) as our goodness-of-fit index. The AIC is suited to comparing models with different degrees of complexity because it penalizes models with additional fit parameters. We calculated raw Akaike values and Akaike Weights (w_i), which are a transformation of raw scores that indicate the probability that a particular model among the set of models is correct (Wagenmakers and Farrell, 2004). Using these weights, we also calculated evidence ratios by dividing the weight of one model by the weight of another. These ratios are understood in context of a "confidence set," which is similar to a confidence interval and is defined as 10% of the highest Akaike Weight in the set (Royall, 1997). For the purposes of interpretation, it should be noted that lower raw Akaike values and higher Akaike Weights indicate a better fit to the data.

As indicated by the evidence ratios in **Table 1**, our curve-fit analyses confirmed that non-linear quadratic and cubic models were best fit to the data, whereas linear models fell outside the confidence set. To the extent that such non-linear functions are a defining feature of the Uncanny Valley (Burleigh et al., 2013), this finding is consistent with the possibility that Uncanny Valley effects can occur with distinctly non-human stimuli.

Indeed, we predicted that the specific shape of this non-linear function for bistable images would further resemble an Uncanny Valley by having lower affective ratings for morph-stimuli near the midpoint of the continua than for those near the endpoints. The average rating for stimuli at positions 1 and 7—the unambiguous category endpoints—was therefore compared to the average rating for the position-4 midpoint stimulus for each stimulus set using paired-samples *t*-tests. Consistent with our expectations, endpoint items were rated more positively than midpoint items from the duck-elephant [endpoints, $M = 0.65$, SD $= 0.17$; midpoint, $M = 0.58$, SD $= 0.22$; $t(59) = 2.86$, $p = 0.006$], rabbit-duck [endpoints, $M = 0.66$, SD $= 0.16$; midpoint, $M = 0.56$, SD $= 0.22$; $t(59) = 4.64$, $p < 0.001$], and rhino-giraffe [endpoints, $M = 0.54$, SD $= 0.17$; midpoint, $M = 0.45$, SD $= 0.18$; $t(59) = 3.80$, $p < 0.001$] sets.

Our results based on average ratings suggest that participants provided their lowest affective ratings to morph-stimuli at intermediate positions of each continuum rather than to those at either continuum endpoint. To examine the extent to which this pattern was observable at the level of individual participants, we plotted an abbreviated rating function for each participant's affective response to each perceptual continuum. This was comprised of each participant's rating of the stimulus at each endpoint (i.e., positions 1 and 7) along with the lowest rating they provided to an intermediate stimulus (i.e., among positions 2–6). As shown in **Figure 3A**, for each perceptual continuum, the vast majority of participants (85% for duck-elephant, 87% for duck-rabbit, 80% for rhino-giraffe) provided their most negative affective rating in response to an intermediate stimulus. The resulting 'valley' shape of these individuals' rating functions is visually evident despite substantial variability otherwise in their individual responses. However, it is also the case that, for each perceptual continuum, a corresponding minority of participants provided their lowest rating for an endpoint stimulus, failing to show a valley shape in their individual rating functions (see **Figure 3B**). This suggests that while most individuals' affective responses were appropriately reflected by our group averages, others do not show the same Uncanny-valley-type pattern of responses (see MacDorman and Entezari, 2015, for another example of individual differences in Uncanny Valley effects).

Taken together, the results of Experiment 1 replicate and expand upon the important preliminary findings of Yamada et al. (2012, 2013) to confirm that Uncanny Valley-type effects can occur with distinctly non-human stimuli. These findings are therefore consistent with the possibility that the drop in affective response reflected by the Uncanny Valley may not be determined by the human-ness of the stimuli *per se*, but might instead occur at whatever point requires greatest inhibition of competing visual-category information to select one stimulus interpretation over another.

Experiment 2: 3D Computer Models

Materials and Methods

The possibility that the Uncanny Valley reflects inhibitory devaluation of stimuli that activate multiple, competing stimulus interpretations during recognition suggests that the location of the

TABLE 1 | Experiment 1 curve fit analyses.

Set	Model	*Residual sum of squares*	*AICc*	Δ_i*(AIC)*	w_i*(AIC)*	*CI*
Duck-elephant*	Linear[1]	20.31	−1270.26	7.66	0.02	0.07
	Quadratic[2]	19.85	−1277.92	0.00	0.71	–
	Cubic[3]	19.84	−1276.03	−1.89	0.28	–
Rabbit-duck*	Linear	19.92	−1278.45	7.62	0.01	0.07
	Quadratic	19.47	−1286.07	0.00	0.65	–
	Cubic	19.44	−1284.71	1.35	0.33	–
Rhino-giraffe*	Linear	20.76	−1261.06	12.91	0.00	0.07
	Quadratic	20.10	−1272.67	1.29	0.34	–
	Cubic	19.94	−1273.96	0.00	0.66	–

[1]*K = 1,* [2]*K = 2,* [3]*K = 3, *n = 420.*

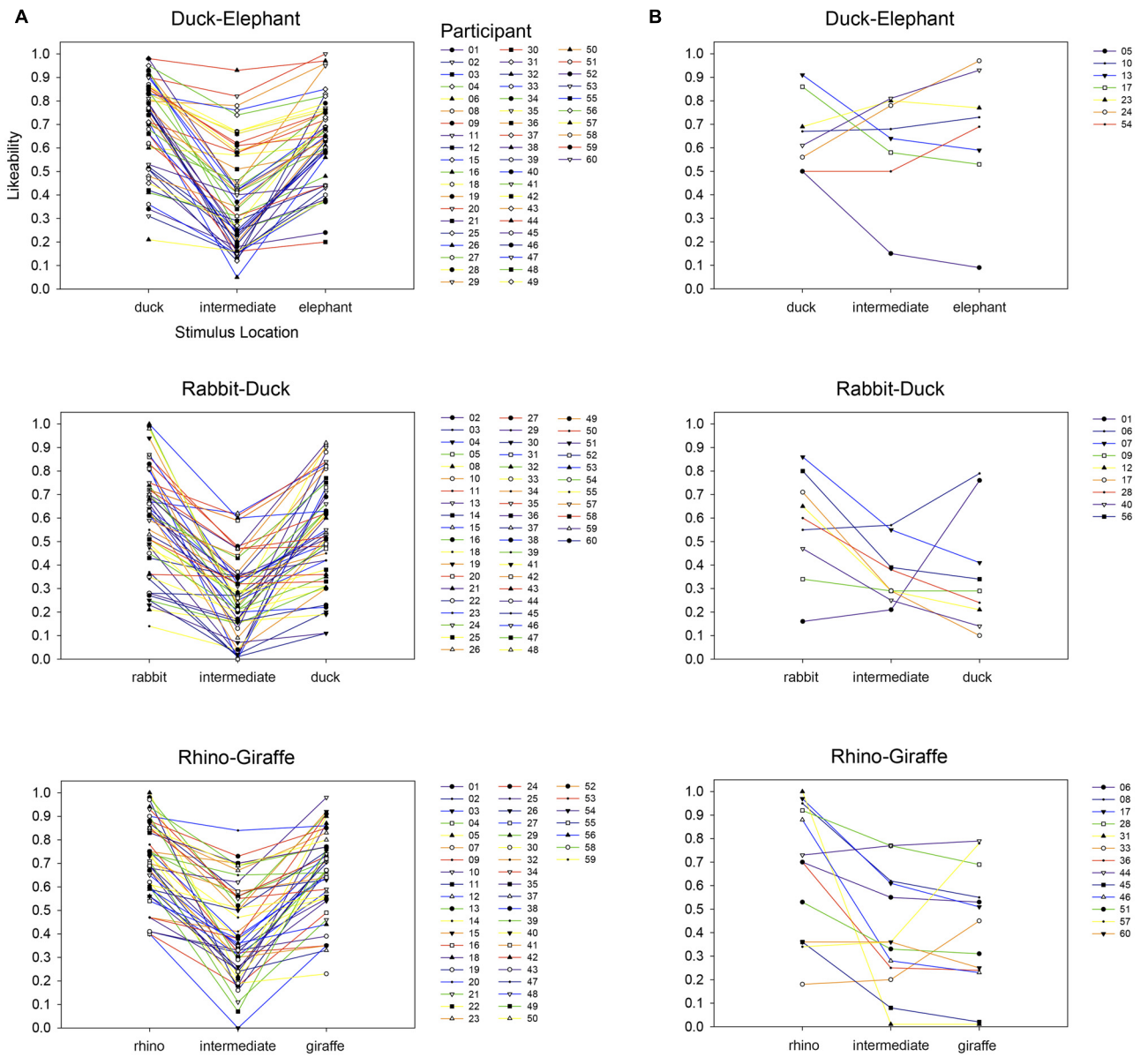


FIGURE 3 | Individual participants' affective response to each animal–animal bistable morph continua in Experiment 1. Abbreviated rating functions are each comprised of a specific participant's rating of the stimulus at each endpoint (i.e., positions 1 and 7) along with the lowest rating they provided to an intermediate stimulus (i.e., among positions 2–6). The majority of participants **(A)** provided their lowest affective rating to a stimulus from an intermediate position. The resulting 'valley' shape of these individual rating functions is consistent with the Uncanny Valley-type effects identified in the group-average results, but is absent in a minority of participants **(B)** who each provided their lowest rating to at least one of the continua-endpoint items.

valley—the continuum point showing the most affectively negative stimulus response— should occur at the point where multiple competing visual-category representations are most strongly activated. Thus, in contrast to explanations that focus on the 'human-ness' of stimuli, this lowest point for human-like stimuli should not always occur on the 'human' side of the perceptual continuum.

To test this prediction, we measured affective responses to a series of 3D computer-modeled morph stimuli representing different locations along different human-non-human continua. To replicate the classic Uncanny Valley effect, one stimulus set was created from human-robot morphs. Additional stimulus sets were created from various human–animal continua. For all stimulus sets, we expected stimuli near continua midpoints to receive more negative affective ratings than those depicting category endpoints. Furthermore, we predicted that the greatest drop in affective ratings would not consistently occur at stimulus-continuum locations near the "human" endpoint, as predicted by

a conspecific pathogen-avoidance account of the Uncanny Valley effect.

All of the following materials and procedures were approved by the Research Ethics Board at the University of Guelph (REB #11NV011).

Participants

Sixty-nine undergraduate students (54 women, $M_{age} = 19$ years, $SD_{age} = 1.2$) participated in exchange for course credit. The only inclusion criterion was having normal or corrected-to-normal vision. None of the participants in Experiment 2 had previously participated in Experiment 1.

Apparatus and stimuli

Stimuli consisted of 35 computer-generated images that were created using Poser (Version 2012, www.smithmicro.com) modeling software and Abrosoft FantaMorph (Version 5.4, www.fantamorph.com) morphing software. In all, there was one human-robot morph-continuum, and four human-animal continua (human-stag, human-tiger, human-lion, and human-bird), each with seven continuum levels – see **Figure 4**.

In order to create the human-animal stimuli, a "base" human model, *Michael 4*, was obtained from daz3d.com, and modified using commercial morph packages for Poser (specifically, the *Leonese* and *Cervus* characters obtained from philosophersegg.com, and the *Bird Cult* character obtained from daz3d.com). Each of these morph packages comprise a predefined set of morphological transformations that can be applied to a base model in order to holistically transform its morphology into the animal character. These packages also contain textures that transform the "skin" of the character into the fur of the animal. A crucial aspect of these morph packages is that the transformations can be applied in a continuous fashion, by assigning values between 0.000 and 1.000 (e.g., a transformation value of 0.500 would be morphologically half-human and half-animal). Similarly, textures can be applied in a continuous fashion, by applying both the human and animal textures to the same figure, and setting them to different levels of opacity (e.g., a 50% opacity overlay would produce a texture that is half-human and half-animal). In order to generate the human-animal morph stimuli, we therefore used morphological transformation and texture opacity values in Poser to create stepwise morphs from one model to another. These morphs represented the following ratios: 0-animal/100-human, 15/85, 30/70, 50/50, 70/30, 85/15, 100/0. In order to create the human-robot morph stimuli, a slightly different approach was taken. Specifically, we used rendered images of Poser models (i.e., the same human model as before, and *KlanK* from daz3d.com), and entered these into FantaMorph software to create a morph sequence. This change was necessary because we were unable to find a suitable robot morph package for Poser that was compatible with the human figure. The human-robot morph stimuli were generated to represent the same ratios as the human-animal stimuli. All stimuli were cropped and saved as JPEG images at a resolution of 912 × 805 pixels.

A pilot study was conducted to ensure that subjective perceptions of the resulting stimuli were consistent with their objective location on the corresponding continuum. In this pilot study, seven participants rated the 35 stimuli, which were presented in randomized order in a single block, on a 7-point Likert scale

FIGURE 4 | Computer-generated morphs (human-robot, human-stag, human-tiger, human-lion, human-bird).

ranging from "human-like" to "animal-like" or from "human-like" to "robot-like." Results indicated that the point of maximal perceptual ambiguity for these stimulus sets was the item within one step of the midpoint (i.e., midpoint plus or minus one step) of each continuum. Response averages indicated that these items were explicitly rated as equally belonging to each of the two categories. Our analyses also revealed that each step in the continuum was perceived as a linear change in the category membership of the model, resulting in overall linear trends for categorical-similarity ratings across each morph continuum. This also ensured that the stimulus at the midpoint of each continuum clearly contained visual-category information from both end-point categories. None of the participants from this pilot-study were utilized in the subsequent affective-rating task.

The testing apparatus and procedures for the affective-rating task were exactly the same as used in Experiment 1, with the exception that participants provided affective ratings for a total of 35 individual stimuli (five sets of seven morphs) in Experiment 2 compared to 21 stimuli in Experiment 1 (three sets of seven drawings).

Results

Average ratings were calculated for each stimulus as a function of its location along its corresponding continuum. These are plotted separately for each stimulus set in **Figure 5**, along with average ratings calculated across the five stimulus sets.

As a formal test for the existence of an Uncanny Valley function, we again used curve fitting analyses. Raw Akaike values, Akaike Weights, and confidence sets were calculated in order to compare the fit of linear, quadratic, and cubic models.

As indicated by the evidence ratios in **Table 2**, our curve-fit analyses confirmed that non-linear quadratic and cubic models were best fit to the data, whereas linear models fell outside the confidence set. To the extent that such non-linear functions are a defining feature of the Uncanny Valley (Burleigh et al., 2013), this finding is indicative of an Uncanny Valley, not only for the human-robot morph continua, but also for each of the human–animal continua.

While the ability to demonstrate an Uncanny Valley for various sets of stimuli is an important manipulation check, our key prediction in Experiment 2 concerned the continuum-location

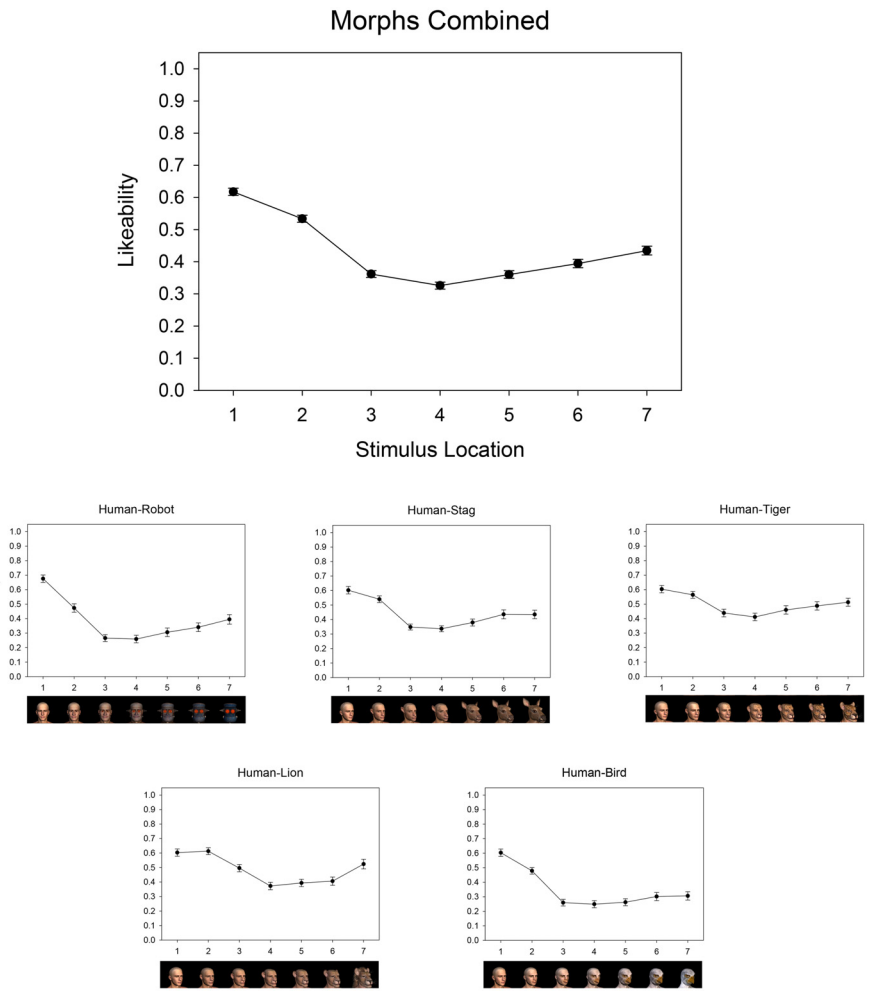


FIGURE 5 | Affective trends for human-robot and human-animal morphs in Experiment 2.

for each stimulus set that received the most negative affective ratings. As shown in **Figure 5**, this valley low-point was located at position 4—the continua midpoint— for each of the human-robot and human-animal stimulus sets (human-robot: $M = 0.26$, SD $= 0.22$; human-stag: $M = 0.33$, SD $= 0.17$; human-tiger: $M = 0.41$, SD $= 0.22$, human-lion: $M = 0.37$, SD $= 0.21$; human-bird: $M = 0.25$, SD $= 0.20$). Moreover, paired-samples t-tests revealed that average ratings for items near these continua midpoints were indeed significantly lower than those for endpoint items for both the human-robot stimuli (endpoints, $M = 0.54$, SD $= 0.17$; midpoint, $M = 0.26$, SD $= 0.22$; $t(68) = 12.95$, $p < 0.001$) and the human–animal stimuli [endpoints, $M = 0.52$, SD $= 0.16$; midpoint, $M = 0.34$, SD $= 0.24$; $t(68) = 9.83$, $p < 0.001$].

As in Experiment 1, we plotted an abbreviated rating function for each participant's affective response to each perceptual continuum comprised of their rating of each endpoint stimulus along with their lowest rating to an intermediate stimulus. As shown in **Figure 6A**, for each continuum, the majority of participants (79% for human-bird, 76% for human-stag, 76% for human-lion, 72% for human-tiger, and 79% for human-robot) provided their most negative affective rating in response to an intermediate stimulus. The resulting 'valley' shape of these individuals' rating functions is visually evident for each continuum. However, a corresponding minority of participants again provided their lowest rating for an endpoint stimulus, failing to show a valley shape in their individual rating functions (see **Figure 6B**). Indeed, certain participants consistently failed to show Uncanny Valley effects for most (e.g., Participant 4) or all (e.g., Participant 11) morph continua.

Taken together, the results of Experiment 2 replicate previous findings (e.g., Burleigh et al., 2013) and are consistent with the possibility that the affective low-points in Uncanny Valley functions are not determined by the human-ness of the stimuli *per se*, but instead by the amount of conflicting stimulus information during recognition and the need to inhibit competing visual-category information to select one stimulus interpretation over another.

Assessing Feasibility: Discussion

The data reported above were collected as a first step in experimentally assessing the feasibility of the hypothesis that the Uncanny Valley reflects a form of inhibitory stimulus devaluation. For a perceptual continuum where the endpoints are comprised of two separate categories, this account predicts that maximum negative affect occurs where there is greatest activation of multiple, competing stimulus representations. We conducted two experiments to test the corresponding predictions (1) that the Uncanny Valley also occurs when humans view distinctly non-human stimuli, and (2) that the lowest point of the 'valley' does not always occur on the 'human' side of the perceptual continuum. In Experiment 1, we used distinctly non-human stimuli: bistable line drawings of animals, a type of stimulus that activates multiple, competing stimulus interpretations. We found that affective ratings of these bistable continua were best fit by non-linear models, which is consistent with an Uncanny Valley interpretation, and that stimuli near the midpoint were rated as least likeable. In Experiment 2, we generated 3D computer-modeled images that comprised human-to-robot and human-to-animal morphed continua. We found that affective ratings of these human-non-human morph continua were best fit by non-linear models, and that stimuli near the midpoint were rated as least likeable. Importantly, we observed that affective minima—the stimulus with the lowest affect rating in each stimulus set—were not always on the 'human' side of the human-non-human continua. Taken together, these results are consistent with a general recognition-related account based on the novel hypothesis that the Uncanny Valley is a specific instance of a more general form of stimulus devaluation that occurs when inhibition is triggered to resolve conflict between competing stimulus-related representations.

TABLE 2 | Experiment 2 curve fit analyses.

Set	Model	*Residual sum of squares*	*AICc*	$\Delta_i(AIC)$	$w_i(AIC)$	*CI*
Human-robot*	Linear[1]	32.69	−1298.74	93.34	0.00	0.09
	Quadratic[2]	27.13	−1386.78	5.29	0.07	–
	Cubic[3]	26.72	−1392.07	0.00	0.93	–
Human-stag*	Linear	23.73	−1453.33	46.15	0.00	0.05
	Quadratic	21.48	−1499.48	0.00	0.47	–
	Cubic	21.38	−1499.74	−0.26	0.53	–
Human-tiger*	Linear	25.78	−1413.40	22.25	0.00	0.03
	Quadratic	24.52	−1435.65	0.00	0.70	–
	Cubic	24.50	−1433.94	1.71	0.30	–
Human-lion*	Linear	26.66	−1452.02	42.92	0.00	0.10
	Quadratic	24.86	−1484.60	10.34	0.01	–
	Cubic	24.25	−1494.94	0.00	0.99	–
Human-bird*	Linear	24.82	−1431.75	67.46	0.00	0.07
	Quadratic	21.58	−1497.26	1.95	0.27	–
	Cubic	21.41	−1499.21	0.00	0.73	–

[1]$K = 1$, [2]$K = 2$, [3]$K = 3$, *$n = 483$.

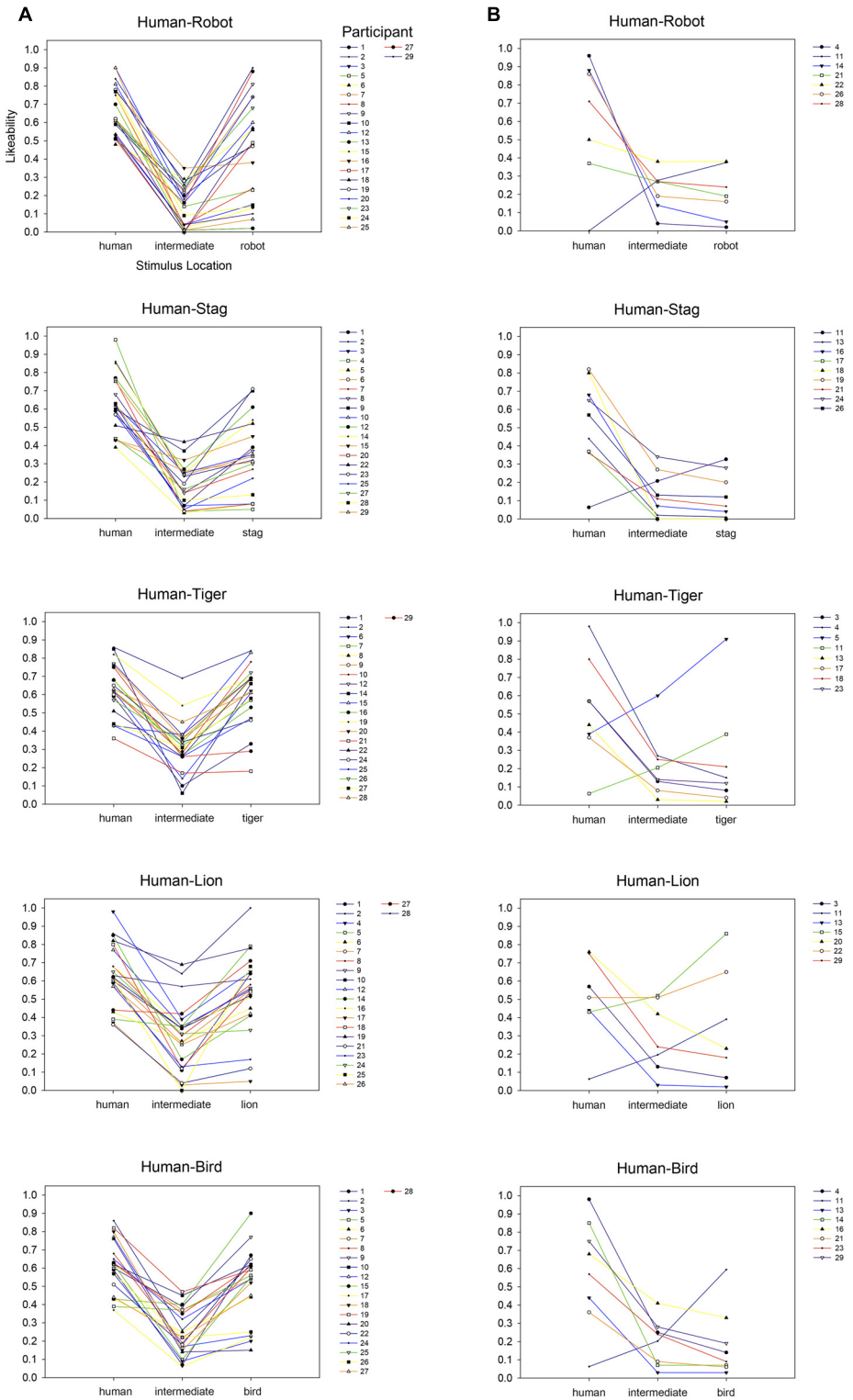


FIGURE 6 | Individual participants' affective response to each human-robot and human-animal morph continua in Experiment 2. Abbreviated rating functions are each comprised of a specific participant's rating of the stimulus at each endpoint (i.e., positions 1 and 7) along with the lowest rating they provided to an intermediate stimulus (i.e., among positions 2–6). The majority of participants **(A)** provided their lowest affective rating to a stimulus from an intermediate position. The resulting 'valley' shape of these individual rating functions is consistent with the Uncanny Valley-type effects identified in the group-average results, but is absent in a minority of participants **(B)** who each provided their lowest rating to at least one of the continua-endpoint items.

Mori's (1970) original hypothesis suggested that near-perfect human-likeness was key to the Uncanny Valley effect, and theoretical accounts such as the pathogen avoidance hypothesis (e.g., MacDorman et al., 2009) were consistent with this premise. These accounts make two basic predictions: the first is that the Uncanny Valley effect should only occur when individuals view stimuli that portray conspecifics (e.g., humans viewing humans), and the second is that the location of the 'valley' should be on the 'human' side of a 'human'/'non-human' continuum. Our results are consistent with a different interpretation. We found an Uncanny Valley effect with distinctly non-human stimuli. Moreover, when using human-like morph stimuli, the location of the 'valley' was not consistently on the 'human' side of the continuum. We suggest instead that the Uncanny Valley may reflect the affective consequences of cognitive processes applied to stimuli whose perception strongly activates multiple, competing stimulus representations. This perspective is based on a well-established and growing body of evidence that cognitive inhibition—known for its role in resolving potential interference during visual tasks—has distinctly negative affective consequences for associated stimuli (Fenske and Raymond, 2006; Frischen et al., 2012). Negative affect for stimuli within an Uncanny Valley context may therefore occur to the extent that selecting one stimulus interpretation over the other requires inhibition of visual-category information associated with the non-selected interpretation. The greater the inhibition during identification, the greater the negative affect for the associated stimulus. Importantly, our results here suggest that such stimulus devaluation characteristic of the Uncanny Valley does not depend on the human-ness of the stimuli *per se*.

Outstanding Issues and Future Directions

Our consideration of prior Uncanny Valley findings and recent advances in our understanding of the affective consequences of cognitive inhibition have led to the hypothesis that affective devaluation by stimulus-related inhibition may underlie the Uncanny Valley effect. The consistency of the results of our preliminary experimental tests of key predictions arising from this new account further establish its feasibility. However, some outstanding issues need to be addressed before the value of inhibitory devaluation can be fully realized as an explanatory construct in Uncanny Valley research.

Negative Affect from Inhibition or Reduced Fluency?

Yamada et al. (2012, 2013) have suggested that the Uncanny Valley effect may be due to low processing fluency (i.e., the ease with which stimulus-information is processed) for items that are difficult to categorize (e.g., as human or non-human). This account relies on the well-established connection between increases in processing fluency and the experience of positive affect toward associated stimuli (see Reber et al., 2004 for review). There are some reasons, however, to suspect that the distinctly negative responses associated with the Uncanny Valley may be linked to cognitive inhibition rather than fluctuations in fluency, *per se*. First, it is conceivable that a stimulus-processing episode involving inhibition might reduce fluency and the positive affect associated with the item—in which case, processing fluency would be a *proxy* for cognitive inhibition. Second, the affective consequences of fluency are thought to be distinctively positive (Reber et al., 1998; Winkielman and Cacioppo, 2001). Furthermore, experimental conditions that typically favor increased perceptual fluency (i.e., repeated and longer-duration stimulus exposures, and stimulus presentations at central foveal locations associated with high visual acuity) nevertheless lead to distinctly negative affective stimulus ratings whenever successful task performance requires attentional or response-related inhibition (Fenske et al., 2004; Raymond et al., 2005; Frischen et al., 2012). Thus, the available evidence to date points to a clearer link between inhibition and aversive stimulus response than between changes in fluency and stimulus-related negative affect.

The Challenge of Indirect Measures of Inhibition

We conducted two experiments to assess the feasibility of the hypothesis that the Uncanny Valley reflects the negative affective consequences of cognitive inhibition, yet neither experiment included a measure of inhibition, *per se*. And while the link between inhibition and stimulus devaluation is sufficiently strong that researchers have begun to take the occurrence of increasingly negative subjective stimulus evaluations as a key indicator of the potential involvement of inhibition at key points within a given task (e.g., Kihara et al., 2011), there are certainly many other factors that can lead to an aversive response. Thus, one of the outstanding challenges for further assessing the feasibility of the inhibitory-devaluation hypothesis is to obtain converging evidence that competition between the multiple stimulus-category representations activated by Uncanny-type stimuli is indeed resolved through inhibition of the non-selected representations. Part of the challenge arises from the fact that traditional cognitive-behavioral measures of inhibition (e.g., perceptual response time and accuracy) are also indirect measures—indices of inhibitory aftereffects rather than a metric of inhibition itself. These traditional measures, as well subjective affective ratings, can therefore be influenced by other factors that may systematically accompany inhibition, such as cognitive conflict. Nevertheless, the absence of a direct behavioral measure of inhibition has not precluded the usefulness of using indirect measures in exploring its potential involvement in a wide variety of cognitive faculties (for review, see Bari and Robbins, 2013). Advances in combining cognitive methods with neuroimaging techniques can also provide a converging-methods approach that may be critical for disentangling the specific cognitive and affective sequence of events involved in inhibitory devaluation and the extent to which they contribute to the Uncanny Valley effect. We outline some of these possibilities below.

Neurocognitive Mechanisms

If the Uncanny Valley effect reflects inhibitory devaluation of stimuli that activate multiple, competing interpretations during recognition, then neuroimaging investigations of the Uncanny Valley should be expected to show critical similarities with neuroimaging investigations of inhibitory devaluation. So far, the only examinations of the neural correlates of inhibitory devaluation include a pair of electrophysiological (event-related potential) studies by Kiss et al. (2007, 2008), and a functional magnetic resonance imaging (fMRI) study by Doallo et al. (2012). Nevertheless, each of these studies have indicated that the magnitude of neural activation associated with resolving potential interference among competing stimulus/motor-response representations are linked to subsequent levels of negative subjective evaluations of the associated stimuli.

Doallo et al. (2012), for example, found that the level of activity in lateral prefrontal cortex (middle frontal gyrus) was greatest during periods requiring response inhibition for successful task performance. The level of activity in this inhibition-related region was linked to the subsequent magnitude of affective devaluation in participants' subjective ratings of the stimuli. Within the realm of Uncanny Valley effects, Saygin et al. (2012) likewise observed a greater change in activity within a region of middle frontal gyrus (among other lateral areas of the parietal and temporal cortices) when participants repeatedly observed a robot with human features (i.e., a stimulus depicting multiple, competing categories) than when participants repeatedly viewed a robot without human features or a real human (i.e., stimuli that depicted a single object category). It should be noted, however, that Saygin et al. (2012) did not measure affective responses to their stimuli. Nevertheless, some broadly consistent findings across paradigms have also been obtained in fronto-limbic areas thought to be involved in the coding of items' motivational or emotional significance. Doallo et al. (2012), for example, reported that the level of activity in orbital-frontal cortex, insular cortex, and amygdala during periods of motor-inhibition was linked to the subsequent magnitude of stimulus devaluation. Cheetham et al. (2011) likewise observed relative increases in activity within amygdala and insular cortex when a morph image bearing a greater resemblance to an inanimate human-like avatar was quickly followed by a different image depicting a real human than when followed by another avatar. Their findings support the idea that conditions that evoke multiple, competing stimulus representations can elicit activity in emotion-related areas, even in the absence of motor-related conflict or categorical ambiguity, *per se*. Unfortunately, their passive-viewing approach meant that participants in their study were not asked to provide any explicit perceptual or (even more importantly here) affective judgments about the stimuli they viewed, making it impossible to link the changes in neural activity they observed to subjective perceptual or affective outcomes. And while comparisons are otherwise limited by the many differences in methods and procedures, the consistency in the results of these prior neuroimaging studies of inhibitory devaluation and Uncanny Valley effects certainly does support a call for future studies to directly examine specific issues regarding the extent to which the Uncanny Valley effect reflects inhibitory devaluation of stimuli that activate multiple, competing stimulus representations.

For example, tasks that involve cognitive conflict are thought to rely on the anterior cingulate cortex for conflict detection (Botvinick et al., 2001; Kerns et al., 2004), while the lateral prefrontal regions appear to be recruited during subsequent cognitive control (Carter and van Veen, 2007). This may explain why tasks that evoke cognitive dissonance have been shown to engage the dACC as well as the insula (van Veen et al., 2009) and left dorsolateral PFC (Mengarelli et al., 2013), and why activation in these areas has predicted attitude change in line with cognitive dissonance theory. Indeed, more recent evidence (Izuma et al., 2013) suggests that the same region of the dACC is involved in both cognitive dissonance and conflict. Therefore, if the Uncanny Valley effect can be understood as the affective consequence of inhibition applied to reduce cognitive conflict, then we might expect to see anterior cingulate involvement when participants judge stimuli that present conflicting cues to category membership. This can be assessed by combining Uncanny Valley paradigms such as the paradigm used in our experiments with techniques such as fMRI or EEG/ERP. Electrophysiological markers of conflict detection, such as the N2 event-related potential obtained from frontocentral electrodes (directly above the dACC), for example, may be particularly useful for examining the link between conflict and stimulus devaluation in Uncanny Valley paradigms.

One possibility would involve extending the approach used by Kiss et al. (2008) in their EEG/ERP study of inhibitory devaluation that focused on how changes in the amplitude of the N2 component were linked to the magnitude of stimulus devaluation measured thereafter. Using this approach with Uncanny Valley-type stimulus sets would likewise be expected to reveal the largest N2 component, and the most negative affective response, for those stimuli that most strongly activate multiple, competing stimulus interpretations. Experimental priming manipulations might also be used to vary the extent to which a stimulus from a given perceptual continuum would activate multiple, competing stimulus interpretations. For target images selected from the midpoints of human-robot morph sequences, for example, varying whether a preceding prime image is either a human, robot, or from a completely unrelated (control) category should have an impact on behavioral and neuroimaging measures of cognitive conflict, inhibition, and subjective emotional responses to the target images. The anterior midcingulate cortex (aMCC) has also been linked to cognitive control and negative affect (Shackman et al., 2011). Future fMRI investigations may therefore target the aMCC as another potential link between cognitive conflict, the recruitment of cognitive inhibition and negative affect in situations involving the Uncanny Valley and other conditions known to produce inhibitory devaluation.

Individual Differences

Another important avenue of research concerns individual differences in inhibitory processes that might explain differing affective responses to 'uncanny' stimuli. In our experimental assessment of the feasibility of the inhibitory-devaluation hypothesis (Experiments 1 and 2),

we observed that, while the individual affective responses of most participants reflected an Uncanny Valley-like pattern of stimulus evaluation, this was not universal. Individual differences in affective responses to uncanny stimuli could arise either due to varying inhibitory control abilities, or due to variations in the affective sensitivity of an individual to an inhibitory signal. For example, MacDorman and Entezari (2015) recently observed that the eeriness induced by uncanny stimuli was more pronounced among individuals with high trait anxiety. Given evidence that trait-anxious persons exhibit hyper-responsivity in fronto-limbic regions associated with negative affect (Shin and Liberzon, 2009), it is possible that their sensitivity to Uncanny stimuli arises due to heightened affective reactivity to inhibitory signals. Individual differences in the magnitude of other forms of inhibitory-devaluation have also been linked to differences in inhibitory control (e.g., failures to inhibit motor-responses, Ferrey et al., 2012). Exploring the relationship between individual differences in the subjective magnitude of the Uncanny Valley effect and differences in inhibitory control and/or affective sensitivity to inhibitory signals may therefore provide another interesting direction for future research into the specific sequence of cognitive and affective events that the inhibitory-devaluation account of the Uncanny Valley is proposed to comprise.

Conclusion

The importance of elucidating specific mechanisms underlying the Uncanny Valley effect is underscored by the extent to which interest in this effect has spread from robotics into other areas, such as computer graphics and prosthetics (Seyama and Nagayama, 2007; MacDorman et al., 2009; Mitchell et al., 2011; Tinwell et al., 2011; Poliakoff et al., 2013). Whereas many accounts of this effect have focused on the potential relationship between the subjective human-likeness of a stimulus and an observer's emotional response to it (e.g., MacDorman and Ishiguro, 2006; Seyama and Nagayama, 2007; MacDorman et al., 2009), we propose an alternate account based on the novel hypothesis that the Uncanny Valley is not directly related to 'human-likeness' *per se,* but instead reflects a more general form of stimulus devaluation that occurs when inhibition is triggered to resolve conflict between competing stimulus-related representations. Our preliminary assessment, both in terms of prior findings as well as through our experimental tests of two specific predictions arising from this new account of the Uncanny Valley, establish its feasibility and make clear the need for additional research into the possibility that the strong dislike of Uncanny-type stimuli reflects the negative affective consequences of conflict-resolving inhibition.

Acknowledgments

This research was supported by the Natural Science and Engineering Research Council of Canada, the Canada Foundation for Innovation, and the Ontario Ministry of Research and Innovation.

References

Bari, A., and Robbins, T. W. (2013). Inhibition and impulsivity: behavioral and neural basis of response control. *Prog. Neurobiol.* 108, 44–79. doi: 10.1016/j.pneurobio

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., and Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychol. Rev.* 108, 624–652. doi: 10.1037/0033-295X.108.3.624

Brehm, J. W. (1956). Postdecision changes in the desirability of alternatives. *J. Abnorm. Soc. Psychol.* 52, 384–389. doi: 10.1037/h0041006

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

Burleigh, T. J., Schoenherr, J. R., and 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. *Comput. Hum. Behav.* 29, 759–771. doi: 10.1016/j.chb.2012.11.021

Burnham, K. P., and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach,* 2nd Edn. New York: Springer.

Carter, C., and van Veen, V. (2007). Anterior cingulate cortex and conflict detection: an update of theory and data. *Cogn. Affect. Behav. Neurosci.* 7, 367–379. doi: 10.3758/CABN.7.4.367

Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., and Jäncke, L. (2013). Category processing and the human likeness dimension of the uncanny valley hypothesis: eye-tracking data. *Front. Psychol.* 4:108. doi: 10.3389/fpsyg.2013.00108

Cheetham, M., Suter, P., and Jäncke, L. (2011). The human likeness dimension of the "uncanny valley hypothesis": behavioral and functional MRI findings. *Front. Hum. Neurosci.* 5:126. doi: 10.3389/fnhum.2011.00126

Cheetham, M., Suter, P., and Jäncke, L. (2014). Perceptual discrimination difficulty and familiarity in the uncanny valley: more like a "happy valley." *Front. Psychol.* 5:1219. doi: 10.3389/fpsyg.2014.01219

Croyle, R. T., and Cooper, J. (1983). Dissonance arousal: physiological evidence. *J. Pers. Soc. Psychol.* 45, 782–791. doi: 10.1037/0022-3514.45.4.782

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222. doi: 10.1146/annurev.ne.18.030195.001205

Doallo, S., Raymond, J. E., Shapiro, K. L., Kiss, M., Eimer, M., and Nobre, A. C. (2012). Response inhibition results in the emotional devaluation of faces: neural correlates as revealed by fMRI. *Soc. Cogn. Affect. Neurosci.* 7, 649–659. doi: 10.1093/scan/nsr031

Elliot, A. J., and Devine, P. G. (1994). On the motivational nature of cognitive dissonance: dissonance as psychological discomfort. *J. Pers. Soc. Psychol.* 67, 382–394. doi: 10.1037/0022-3514.67.3.382

Fenske, M. J., and Raymond, J. E. (2006). Affective influences of selective attention. *Curr. Dir. Psychol. Sci.* 15, 312–316. doi: 10.1111/j.1467-8721.2006.00459.x

Fenske, M. J., Raymond, J. E., Kessler, K., Westoby, N., and Tipper, S. P. (2005). Attentional inhibition has social-emotional consequences for unfamiliar faces. *Psychol. Sci.* 16, 753–758. doi: 10.1111/j.1467-9280.2005.01609.x

Fenske, M. J., Raymond, J. E., and Kunar, M. A. (2004). The affective consequences of visual attention in preview search. *Psychon. Bull. Rev.* 11, 1034–1040. doi: 10.3758/BF03196736

Ferrey, A. E., Frischen, A., and Fenske, M. J. (2012). Hot or not: response inhibition reduces the hedonic value and motivational incentive of sexual stimuli. *Front. Psychol.* 3:575. doi: 10.3389/fpsyg.2012.00575

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.

Fisher, G. H. (1967). Measuring ambiguity. *Am. J. Psychol.* 53, 541–557. doi: 10.2307/1421187

Frischen, A., Ferrey, A. E., Burt, D. H. R., Pistchik, M., and Fenske, M. J. (2012). The affective consequences of cognitive inhibition: devaluation or neutralization? *J. Exp. Psychol. Hum. Percept. Perform.* 38, 169–179. doi: 10.1037/a0025981

Griffiths, O., and Mitchell, C. J. (2008). Negative priming reduces affective ratings. *Cogn. Emot.* 22, 1119–1129. doi: 10.1080/02699930701664930

Harmon-Jones, E. (2004). Contributions from research on anger and cognitive dissonance to understanding the motivational functions of asymmetrical frontal brain activity. *Biol. Psychol.* 67, 51–76. doi: 10.1016/j.biopsycho.2004.03.003

Hogeboom, M. M. (1995). *On the Dynamics of Static Pattern Perception*. Amsterdam: University of Amsterdam.

Izuma, K., Matsumoto, M., Murayama, K., Samejima, K., Norihiro, S., and Matsumoto, K. (2013). "Neural correlates of cognitive dissonance and decision conflict," in *Advances in Cognitive Neurodynamics (III)*, ed. Y. Yamaguchi (Dordrecht: Springer Science and Business Media), 623–628.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., and Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science* 303, 1023–1026. doi: 10.1126/science.1089910

Kihara, K., Yagi, Y., Takeda, Y., and Kawahara, J. I. (2011). Distractor devaluation effect in the attentional blink: direct evidence for distractor inhibition. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 168–179. doi: 10.1037/a0019948

Kiss, M., Goolsby, B. A., Raymond, J. E., Shapiro, K. L., Silvert, L., Nobre, A. C., et al. (2007). Efficient attentional selection predicts distractor devaluation: event-related potential evidence for a direct link between attention and emotion. *J. Cogn. Neurosci.* 19, 1316–1322. doi: 10.1162/jocn.2007.19.8.1316

Kiss, M., Raymond, J. E., Westoby, N., Nobre, A. C., and Eimer, M. (2008). Response inhibition is linked to emotional devaluation: behavioural and electrophysiological evidence. *Front. Hum. Neurosci.* 2:13. doi: 10.3389/neuro.09.013.2008

Losch, M. E., and Cacioppo, J. T. (1990). Cognitive dissonance may enhance sympathetic tonus, but attitudes are changed to reduce negative affect rather than arousal. *J. Exp. Soc. Psychol.* 26, 289–304. doi: 10.1016/0022-1031(90)90040-S

MacDorman, K. F., and Entezari, S. O. (2015). Individual differences predict sensitivity to the uncanny valley. *Interact. Stud.* IS-D-13-00026R2.

MacDorman, K. F., Green, R. D., Ho, C.-C., and Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Comput. Hum. Behav.* 25, 695–710. doi: 10.1016/j.chb.2008.12.026

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

Martiny-Huenger, T., Gollwitzer, P. M., and Oettingen, G. (2013). Distractor devaluation in a flanker task: object-specific effects without distractor recognition memory. *J. Exp. Psychol. Hum. Percept. Perform.* 40, 613–625. doi: 10.1037/a0034130

Meng, M., and Tong, F. (2004). Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *J. Vis.* 4, 539–551. doi: 10.1167/4.7.2

Mengarelli, F., Spoglianti, S., Avenanti, A., and Di Pellegrino, G. (2013). Cathodal tDCS over the left prefrontal cortex diminishes choice-induced preference change. doi: 10.1093/cercor/bht314 [Epub ahead of print].

Mitchell, W. J., Szerszen, K. A. Sr., Lu, A. S., Schermerhorn, P. W., Scheutz, M., and MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *Iperception* 2, 10–12. doi: 10.1068/i0415

Moore, R. K. (2012). A Bayesian explanation of the "uncanny valley" effect and related psychological phenomena. *Sci. Rep.* 2:864. doi: 10.1038/srep00864

Mori, M. (1970). Bukimi no tani [the Uncanny Valley]. *Energy* 7, 33–35.

Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., and O'Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends Cogn. Sci.* 15, 453–459. doi: 10.1016/j.tics.2011.07.011

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *J. Neurosci. Methods* 162, 8–13. doi: 10.1016/j.jneumeth.2006.11.017

Poliakoff, E., Beach, N., Best, R., Howard, T., and Gowen, E. (2013). Can looking at a hand make your skin crawl? Peering into the uncanny valley for hands. *Perception* 42, 998–1000. doi: 10.1068/p7569

Raymond, J. (2009). Interactions of attention, emotion and motivation. *Prog. Brain Res.* 176, 293–308. doi: 10.1016/S0079-6123(09)17617-3

Raymond, J. E., Fenske, M. J., and Tavassoli, N. T. (2003). Selective attention determines emotional responses to novel visual stimuli. *Psychol. Sci.* 14, 537–542. doi: 10.1046/j.0956-7976.2003.psci_1462.x

Raymond, J. E., Fenske, M. J., and Westoby, N. (2005). Emotional devaluation of distracting stimuli: a consequence of attentional inhibition during visual search? *J. Exp. Psychol. Hum. Percept. Perform.* 31, 1404–1415. doi: 10.1037/0096-1523.31.6.1404

Reber, R., Schwarz, N., and Winkielman, P. (2004). Processing fluency and aesthetic pleasure: is beauty in the perceiver's processing experience? *Pers. Soc. Psychol. Rev.* 8, 364–382. doi: 10.1207/s15327957pspr0804_3

Reber, R., Winkielman, P., and Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychol. Sci.* 9, 45–48. doi: 10.1111/1467-9280.00008

Royall, R. M. (1997). *Statistical Evidence: A Likelihood Paradigm*. New York: Chapman & Hall.

Rozin, P., and Fallon, A. E. (1987). A perspective on disgust. *Psychol. Rev.* 94, 23–41. doi: 10.1037/0033-295X.94.1.23

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., and Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* 7, 413–422. doi: 10.1093/scan/nsr025

Seyama, J., and Nagayama, R. S. (2007). The uncanny valley: effect of realism on the impression of artificial human faces. *Presence (Cambridge, MA)* 16, 337–351. doi: 10.1162/pres.16.4.337

Shackman, A. J., Salomons, T. V., Slagter, H., Fox, A. S., Winter, J. J., and Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat. Rev. Neurosci.* 12, 154–167. doi: 10.1038/nrn2994

Shin, L. M., and Liberzon, I. (2009). The neurocircuitry of fear, stress, and anxiety disorders. *Neuropsychopharmacology* 35, 169–191. doi: 10.1038/npp.2009.83

Sterzer, P., Kleinschmidt, A., and Rees, G. (2009). The neural bases of multistable perception. *Trends Cogn. Sci.* 13, 310–318. doi: 10.1016/j.tics.2009.04.006

Stroop, J. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662. doi: 10.1037/h0054651

Tinwell, A., Grimshaw, M., Nabi, D. A., and Williams, A. (2011). Facial expression of emotion and perception of the uncanny valley in virtual characters. *Comput. Hum. Behav.* 27, 741–749. doi: 10.1016/j.chb.2010.10.018

van Veen, V., Krug, M. K., Schooler, J. W., and Carter, C. S. (2009). Neural activity predicts attitude change in cognitive dissonance. *Nat. Neurosci.* 12, 1469–1474. doi: 10.1038/nn.2413

Verstijnen, I. M., and Wagemans, J. (2004). Ambiguous figures: living versus nonliving objects. *Perception* 33, 531–546. doi: 10.1068/p5213

Wagenmakers, E.-J., and Farrell, S. (2004). AIC model selection using Akaike weights. *Psychon. Bull. Rev.* 11, 192–196. doi: 10.3758/BF03206482

Winkielman, P., and Cacioppo, J. T. (2001). Mind at ease puts a smile on the face: psychophysiological evidence that processing facilitation elicits positive affect. *J. Pers. Soc. Psychol.* 81, 989–1000. doi: 10.1037/0022-3514.81.6.989

Yamada, Y., Kawabe, T., and Ihaya, K. (2012). Can you eat it? A link between categorization difficulty and food likability. *Adv. Cogn. Psychol.* 8, 248–254. doi: 10.5709/acp-0120-2

Yamada, Y., Kawabe, T., and Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *Jpn. Psychol. Res.* 55, 20–32. doi: 10.1111/j.1468-5884.2012.00538.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in PSYCHOLOGY

OPINION ARTICLE
published: 21 January 2015
doi: 10.3389/fpsyg.2014.01456

Uncanny sociocultural categories

Jordan R. Schoenherr[1]* and Tyler J. Burleigh[2]

[1] Department of Psychology, Carleton University, Ottawa, ON, Canada
[2] Department of Psychology, University of Guelph, Guelph, ON, Canada
*Correspondence: jordan.schoenherr@carleton.ca

Edited by:
Marcus Cheetham, University of Zürich, Switzerland

Reviewed by:
Steven Young, Baruch College, USA

Keywords: uncanny valley, cultural artifacts, categorical perception, categorization, social identity, cultural transmission

INTRODUCTION

Humans are well adapted to their social environments. Experimental evidence suggests that humans are either born with, or quickly learn, the necessary affective and cognitive processes that allow them to recognize others, and to understand their mental states and social behavior. Mori's (1970) proposal of an uncanny valley, which describes affective response as a function of distance from a human category defined by morphological and behavioral features (i.e., human likeness), appears to be a sensible extension of these ideas. Following Mori's initial proposal, the uncanny valley has largely been considered in the context of cultural artifacts such as robotics, prosthetics, toys, and puppets. He associated "healthy people" with the greatest level of familiarity and positive affect, prosthetic hands and corpses with a global negative affective maximum, and bunraku puppets and humanoid robots with intermediate levels of familiarity and positive affect. It is important to note that these cultural artifacts represent the most contemporary features of human societies. The uncanny valley likely depends on extensions of prepotent responses to stimuli via general learning mechanisms (e.g., face recognition; Haxby et al., 2002; Sperber and Hirschfeld, 2004). Empirical studies of the uncanny valley have just begun to explore the authenticity of Mori's proposal.

Contemporary studies examining the uncanny valley hypothesis have drawn heavily on the psychological literature to explain these phenomena. The shift from an account of the uncanny valley based on a dimension of human-likeness to that of categorization and frequency-based exposure (Burleigh and Schoenherr, 2014) suggests that classes of cultural artifacts might provide evidence for the ubiquity of phenomena across cultures and within human history. These social representations can become external representations available to all members of a human group and can thereby increase familiarity and anchor human judgments (Moscovici, 1981). Supporting Mori's initial claim, negative responses are the result of a lack of familiarity (e.g., Burleigh and Schoenherr, 2014) that emerge over the course of development (Lewkowicz and Ghazanfar, 2012) as humans' affective systems have yet to adapt to these artifacts. If the uncanny valley does have a general cognitive basis, then evidence from affective, behavioral, and cognitive paradigms should exist both across cultures as well as within human history. These social representations will consequently affect observers' judgments.

HUMAN AND NON-HUMAN CLASSIFICATION

Folktaxonomies and covert categories

A particularly compelling source of evidence for the uncanny valley comes from research into folktaxonomies. When we encounter an organism, our knowledge of folkbiological categories can cause us to classify stimuli in terms of a species (e.g., "fish") or an ecological niche (e.g., "aquatic habitat") that is available within a folktaxonomic structure. While the preferred level of categorization within these taxonomies differs between cultures (e.g., Rosch et al., 1976; Medin et al., 1997) and expertise (Tanaka and Taylor, 1991; Medin et al., 1997), such taxonomies form the basis for all judgements of category membership. In the context of this work, we suggest that cognitive anthropological research on folktaxonomies has revealed uncanny valley-like phenomena in the form of "covert categories"–categories that cannot be readily placed into a taxonomical structure (e.g., octopus). Covert categories are cognitively isolated from other ontological categories (Berlin, 1974; Atran, 1983). For instance, informants might be able to identify a number of basic-level properties of an octopus, and yet be unable to associate it with a superordinate category (e.g., "fish"). In Berlin et al.'s (1968) study of Tzetal Mayans' folktaxonomies, they found numerous covert groups that did not fall into any of the major life-form categories (this term is used in the anthropological literature but is equivalent to the superordinate level in the psychological literature; c.f., Brown, 1974). These categories are also associated with aversive responses, such as food prohibitions (Douglas, 1966/2002; Sperber, 1996). For example, Henrich and Henrich (2010) observed that the ambiguity in classifying an octopus as a "fish" or "non-fish" was associated with a food taboo. Douglas (1957) also observed similar outcomes with other ambiguous animals, like the flying squirrel. Such responses to categorically ambiguous stimuli are consistent with the uncanny valley hypothesis.

Anomalies: gods and monsters

Related to covert categories is the human concern with biological anomalies, gods, and monsters evidenced throughout human history. In many cases, covert categories might be the basis for these ontological categories. As Sperber (1996)

has pointed out in his discussion of hybrids and monsters, such entities tend to combine features from disparate categories; ranging from the addition of a single feature, such as *unicorns* which are horses with a horn, to more elaborate amalgams that combine multiple features, such as the *Minotaur* and *Manticore* which share features from human and animal categories. Representative of the cross-cultural trend are Japanese mythological creatures such as *Tsuchigumo*, a creature that has the face of a demon, the legs of a spider, and the body of a tiger.

Such mythological creatures might reflect social representations that are the products of cultural transmission. This is evidenced by Mayor's (2001, 2005) study of how ancient peoples classified fossilized creatures. Specifically, Mayor noted that ancient peoples appear to have understood fossils by assimilating them into extant ontological categories. In the absence of a clear understanding the process of fossilization, they would have perceived fossils as the scattered bones of what were believed to be living species. For instance, Mayor (2001) claims that when the fossilized remains of a protoceratops were encountered, in an attempt to account for the presence of these bones, ancients inferred the existence of griffons. Support for her argument is taken from the co-localization of fossils and reported observation of these creatures in ancient times. To ancients, these mythical creatures represented fearful stimuli that were associated with distant unknown territories inhabited by uncanny creatures.

Whereas fossil remains and unexplained natural events might help to explain the origin of mythological creatures (Mayor, 2001; Kaplan, 2012), along with responses such as fear and anxiety (also see Asma, 2009), the mechanisms that support the retention of these creatures in the modern mind requires further explanation (e.g., Barrett and Keil, 1996). Similar to the uncanny valley studies investigating categorization, studies of religious beliefs assume that belief in such ontological categories is a result of humans using existing cognitive templates that have exceptional features (Boyer, 1993, 2001). A heightened sensitivity to these anomalous stimuli, or so-called "counterintuitive beliefs" (Boyer and Ramble, 2001; Atran and Norenzayan, 2004), could be a result of increases in negative affect associated with the uncanny valley. These results can be related back to the idea of distinctiveness in memory (e.g., Hunt and Worthen, 2006), wherein the dissimilarity of an item within a given context facilitates the encoding and retrieval of stimuli. The uncanny valley could also have implications for recall that facilitates the cultural transmission of knowledge. As in the case of food taboos, such a negative valence might reduce our willingness to interact with features of our environment, thereby further reducing our exposure to a range of stimuli. An item's distinctiveness in memory can thereby compound an initial aversive response. Modern equivalents are also evidenced in genetically modified organisms that antagonist interest groups have labeled "Frankenfoods," such as the transgenic tomato, which has been engineered with a gene from the winter flounder that makes it tolerant to freezing temperatures. Such rhetorical devices are clearly predicated on a fear of novel hybrid organisms.

Human categories and out-group bias

A final form of cross-cultural evidence for uncanny valley-like phenomena is the perception of human groups. Much like some non-human categories, unfamiliar human groups might be construed as distinct species. For instance, Gil-White (2001) suggests that this could in fact be the case for ethnicities. Identifying a race as "sub-human" implicitly or explicitly can be understood in these terms. For instance, a low-frequency of exposure to out-group members, or explicitly transmitted out-group biases, could create negative affective responses to features associated with these individuals (for the effect of stimulus frequency on affect, see Zajonc, 1968; Bornstein, 1989). Though controversial, numerous studies have found evidence for implicit negative associations with minority groups (Greenwald et al., 2009), which can be contrasted against explicit biases to "lower" social classes, and castes associated with "untouchability," contamination, and food taboos (Harper, 1964). Accounts of early explorers encountering new tribes and peoples for the first time are also consistent with this possibility (e.g., Hall, 1992). Prophylactic measures were taken when entering strange lands inhabited by others and ritual purification following contact was needed to "guard against... the magical arts of its inhabitants" (Frazer, 1922, p. 110; see also Douglas, 1966/2002). For these explorers, the people that they encountered were similar to them but the comparatively small differences in terms of physical and cultural variation produced strong negative affective responses. As Hall (1992) notes of travelers' tales, features of out-group members were sometimes even perceived to reflect a blending of human and non-human animal traits: "In the land of Indian there are men with dogs' heads." (quoted in Newby, 1975, p. 17).

A large degree of variation is also observed in the number and nature of reified gender and sexual categories across societies. Early First Nations societies in North America often recognized three or more sexes or genders and their associated roles within the society (Herdt, 1994). In sharp contrast, the early definition of homosexuality as a disorder in North American psychiatry reflects a perceived "deviation" from socially defined categories (Zucker, 2005). In the context of the present review, the experiences of homosexual, bisexual, and transgendered persons could reflect their status as a covert category in contemporary North American society. For instance, as described in a report by the San Francisco Human Rights Commission (2011), the experiences of bisexuals include being "rendered invisible" and being seen as vectors for the spread of sexually-transmitted diseases. Similar claims could be made concerning the historic status of women in Western and Near Eastern societies. For example, the unitary gender structure in the mythology of Abrahamic religions that sees Eve created from Adam's rib, framed the female sex as a derivative of the male sex. When framed in terms of deviation from a male reference category, the uncanny valley might offer a plausible basis for explaining negative affect and discrimination toward women. While exposure to male and female exemplars should be present in a society with nearly equal frequency, sociocultural practices can limit

the exposure that one sex and gender has to another. While frequency alone is unlikely to account for all sex and gender biases, inasmuch as it does make a contribution it would support the existence of an uncanny valley. An understanding of the conditions in which the uncanny valley occurs, and whether increased exposure to low-frequency or negative-valence categories buffers against it, could facilitate our understanding of intergroup conflicts, and how they can be minimized. The codification of third or multiple gender categories minimally suggests that this is possible.

CONCLUDING REMARKS

Considered individually, folkbiological categories, biological anomalies and monsters, as well as human categories represent individual cultural products of human categorization. Instead, we suggest that the uncanny valley might reflect a primary response to unfamiliar or covert categories. In the absence of having prior knowledge of an individual or group, the relative distinctiveness of a category, due to a lower frequency of exposure, will produce negative affect—an inversion of the mere-exposure effect. The deceptive simplicity of learning mechanisms can lead to important individual and social consequences. If the uncanny valley and its relation to negative affect is a result of frequency of exposure, then its amelioration can be facilitated by increasing the frequency of the target stimuli within the environment. The few studies that have considered the relationship between familiarity, discriminability, and affect (Cheetham et al., 2013, 2014; Burleigh and Schoenherr, 2014) need to be complimented with more research that systematically manipulates the features of the ontological categories used for comparison.

REFERENCES

Asma, S. T. (2009). *On Monsters: An Unnatural History of our Worst Fears*. Oxford: Oxford University Press.

Atran, S. (1983). Covert fragmenta and the origins of the botanical family. *Man* 18, 51–71. doi: 10.2307/2801764

Atran, S., and Norenzayan, A. (2004). Religion's evolutionary landscape: counterintuition, commitment, compassion, communion. *Behav. Brain Sci.* 27, 713–770. doi: 10.1017/S0140525X04000172

Barrett, J. L., and Keil, F. C. (1996). Conceptualizing a nonnatural entity: anthropomorphism in god concepts. *Cogn. Psychol.* 31, 219–247. doi: 10.1006/cogp.1996.0017

Berlin, B. (1974). Further notes on covert categories and folk taxonomies: a reply to Brown. *Am. Anthropol.* 76, 327–331. doi: 10.1525/aa.1974.76.2.02a00080

Berlin, B., Breedlove, D. E., and Raven, P. H. (1968). Covert categories and folk taxonomies, *Am. Anthropol.* 70, 290–299. doi: 10.1525/aa.1968.70.2.02a00050

Bornstein, R. F. (1989). Exposure and affect: overview and meta-analysis of research, 1968–1987. *Psychol. Bull.* 106, 265–289. doi: 10.1037/0033-2909.106.2.265

Boyer, P. (1993). *Cognitive Aspects of Religious Symbolism*. Cambridge: Cambridge University Press.

Boyer, P. (2001). *Religion Explained: The Evolutionary Origins of Religious Thought*. New York, NY: Basic Books.

Boyer, P., and Ramble, C. (2001). Cognitive templates for religious concepts: cross-cultural evidence for recall of counter-intuitive representations. *Cogn. Sci.* 25, 535–564. doi: 10.1207/s15516709cog2504_2

Brown, C. H. (1974). Unique beginners and covert categories in folk biological taxonomies. *Am. Anthropol.* 76, 325–327. doi: 10.1525/aa.1974.76.2.02a00070

Burleigh, T. J., and Schoenherr, J. R. (2014). A reappraisal of the Uncanny Valley: categorical perception or frequency-based sensitization? *Front. Psychol.* 5:1488. doi: 10.3389/fpsyg.2014.01488

Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., and Jancke, L. (2013). Category processing and the humanlikeness dimension of the Uncanny Valley Hypothesis: eye-tracking data. *Front. Psychol.* 4:108. doi: 10.3389/fpsyg.2013. 00108

Cheetham, M., Suter, P., and Jancke, L. (2014). Perceptual discrimination difficulty and familiarity in the Uncanny Valley: more like a "Happy Valley." *Front. Psychol.* 5:1219. doi: 10.3389/fpsyg.2013. 01219

Douglas, M. (1957). Animals in Lele religious symbolism. *Afr. J. Int. Afr. Inst.* 27, 46–58. doi: 10.2307/1156365

Douglas, M. (1966/2002). *Purity and Danger: An Analysis of Concepts of Pollution and Taboo*. London: Routledge & Kegan Paul. doi: 10.4324/9780203361832

Frazer, J. G. (1922). *The Golden Bough: A Study in Magic and Religion*. New York, NY: Macmillan. (Original work published 1890).

Gil-White, F. (2001). Are ethnic groups biological "species" to the human brain? *Curr. Anthropol.* 42, 515–554. doi: 10.1086/321802

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., and Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *J. Pers. Soc. Psychol.* 97, 17–41. doi: 10.1037/a0015575

Hall, S. (1992). "The west and the rest: discourse in power," in *The Indigenous Experience: Global Perspectives*, eds R. C. A. Maaka and C. Andersen (Toronto: Canadian Scholars' Press Inc.), 165–173.

Harper, E. B. (1964). Ritual pollution as an integrator of caste and religion. *J. Asian Stud.* 23, 151–197. doi: 10.2307/2050627

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biol. Psychiatry* 51, 59–67. doi: 10.1016/S0006-3223(01)01330-0

Henrich, J., and Henrich, N. (2010). The evolution of cultural adaptations: Fijian food taboos protect against dangerous marine toxins. *Proc. R. Soc. B* 277, 3715–3724. doi: 10.1098/rspb.2010.1191

Herdt, G., (ed.). (1994). *Third Sex, Third Gender: Beyond Sexual Dimorphism in Culture and History*. New York: Zone Books.

Hunt, R. R., and Worthen, J. (2006). *Distinctiveness and Memory*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195169669.001.0001

Kaplan, M. (2012). *Medusa's Gaze and Vampire's Bite: The Science of Monsters*. New York, NY: Scribner.

Lewkowicz, D. J., and Ghazanfar, A. (2012). The development of the uncanny valley in infants. *Dev. Psychobiol.* 54, 124–32. doi: 10.1002/dev.20583

Mayor, A. (2001). *The First Fossil Hunters: Paleontology in Greek and Roman Times*. Princeton: Princeton University Press.

Mayor, A. (2005). *Fossil Legends of the First Americans*. Princeton, NJ: Princeton University Press.

Medin, D. L., Lynch, E. B., Coley, J. D., and Atran, S. (1997). Categorization and reasoning among tree experts: do all roads lead to Rome? *Cogn. Psychol.* 32, 49–96. doi: 10.1006/cogp.1997.0645

Mori, M. (1970). The uncanny valley. *Energy* 7, 33–35.

Moscovici, S. (1981). "On social representations," in *Social Cognition: Perspectives on Everyday Understanding*, ed J. P. Forgas (London: Academic Press), 181–209.

Newby, E. (1975). *The Mitchell Beazley World Atlas of Exploration*. London: Mitchell Beazley.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., and Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognit. Psychol.* 8, 382–439.

San Francisco Human Rights Commission. (2011). *Bisexual Invisibility: Impacts and Recommendations*. Available online at: http://sf-hrc.org/sites/sf-hrc.org/files/migrated/FileCenter/Documents/HRC_Publications/Articles/Bisexual_Invisiblity_Impacts_and_Recommendations_March_2011.pdf

Sperber, D. (1996). Why are perfect animals, hybrids, and monsters food for symbolic thought? *Method Theory Study Relig.* 8, 143–169. doi: 10.1163/157006896X00170

Sperber, D., and Hirschfeld, L. A. (2004). The cognitive foundations of cultural stability and diversity. *Trends Cogn. Sci.* 8, 40–46. doi: 10.1016/j.tics.2003.11.002

Tanaka, J., and Taylor, M. (1991). Object categories and expertise: is the basic level in the eye of the beholder? *Cognit. Psychol.* 23, 457–482.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *J. Pers. Soc. Psychol.* 9, 1–27. doi: 10.1037/h0025848

Zucker, K. J. (2005). Was the Gender Identity Disorder of childhood diagnosis introduced into DSM-III as a backdoor maneuver to replace homosexuality? A historical note. *J. Sex Marital Ther.* 31, 31–42. doi: 10.1080/00926230590475251

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 10 June 2014; accepted: 28 November 2014; published online: 21 January 2015.

*Citation: Schoenherr JR and Burleigh TJ (2015) Uncanny sociocultural categories. Front. Psychol. **5**:1456. doi: 10.3389/fpsyg.2014.01456*

This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.

frontiers in Psychology

ORIGINAL RESEARCH
published: 15 July 2015
doi: 10.3389/fpsyg.2015.00981

Arousal, valence, and the uncanny valley: psychophysiological and self-report findings

Marcus Cheetham[1,2†], Lingdan Wu[3,4†], Paul Pauli[4] and Lutz Jancke[1]*

[1] Department of Neuropsychology, University of Zurich, Zurich, Switzerland, [2] Department of Psychology, Nungin University, Seoul, South Korea, [3] Swiss Centre for Affective Sciences, University of Geneva, Geneva, Switzerland, [4] Department of Psychology, University of Wurzburg, Wurzburg, Germany

The main prediction of the *Uncanny Valley Hypothesis* (*UVH*) is that observation of humanlike characters that are difficult to distinguish from the human counterpart will evoke a state of negative affect. Well-established electrophysiological [*late positive potential* (*LPP*) and facial *electromyography* (EMG)] and self-report [*Self-Assessment Manikin (SAM)*] indices of valence and arousal, i.e., the primary orthogonal dimensions of affective experience, were used to test this prediction by examining affective experience in response to categorically ambiguous compared with unambiguous avatar and human faces ($N = 30$). LPP and EMG provided direct psychophysiological indices of affective state during passive observation and the SAM provided self-reported indices of affective state during explicit cognitive evaluation of static facial stimuli. The faces were drawn from well-controlled morph continua representing the UVH' *dimension of human likeness* (*DHL*). The results provide no support for the notion that category ambiguity along the DHL is specifically associated with enhanced experience of negative affect. On the contrary, the LPP and SAM-based measures of arousal and valence indicated a general increase in negative affective state (i.e., enhanced arousal and negative valence) with greater morph distance from the human end of the DHL. A second sample ($N = 30$) produced the same finding, using an *ad hoc* self-rating scale of feelings of familiarity, i.e., an oft-used measure of affective experience along *the UVH' familiarity* dimension. In conclusion, this multi-method approach using well-validated psychophysiological and self-rating indices of arousal and valence rejects – for passive observation and for explicit affective evaluation of static faces – the main prediction of the UVH.

Keywords: valence, arousal, uncanny valley hypothesis, familiarity, EMG, EEG, LPP

Edited by:
Eddy J. Davelaar, Birkbeck, University of London, UK

Reviewed by:
Francesca M. M. Citron, Lancaster University, UK
Christian Becker-Asano, Albert-Ludwigs-Universität Freiburg, Germany

***Correspondence:**
Marcus Cheetham, Department of Neuropsychology, University of Zurich, Binzmühlestrasse 14/Box 25, CH-8050 Zürich, Switzerland
m.cheetham@psychologie.uzh.ch

† These authors share first authorship.

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *15 November 2014*
Accepted: *29 June 2015*
Published: *15 July 2015*

Citation:
Cheetham M, Wu L, Pauli P and Jancke L (2015) Arousal, valence, and the uncanny valley: psychophysiological and self-report findings. Front. Psychol. 6:981. doi: 10.3389/fpsyg.2015.00981

Introduction

The longstanding *Uncanny Valley Hypothesis* (*UVH*) predicts that difficulty distinguishing a realistic humanlike character or object (e.g., robot, prosthetic hand) from its human counterpart will evoke an unpleasant affective state (Mori, 1970; **Figure 1**). Mori suggests that this state is characterized by a sense of strangeness and personal disquiet and, when experienced more intensely, by revulsion and disgust. Attention to the originally untested UVH has been spurred

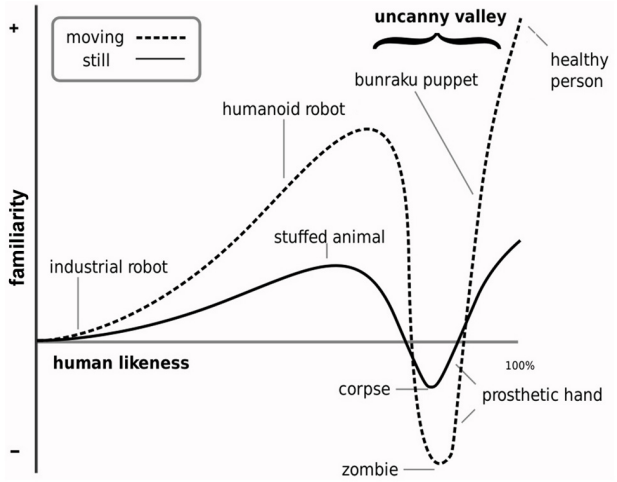


FIGURE 1 | Illustration of the Uncanny Valley Hypothesis. The main prediction of the Uncanny Valley Hypothesis is that observation of highly humanlike characters and objects (depicted along the dimension of human likeness) will evoke a sharp negative peak in affective experience (depicted along the familiarity dimension). This negative peak, referred to as the uncanny valley, is characterized by feelings of strangeness and disquiet. These feelings are suggested to be stronger for dynamic stimuli. The valley occurs at the point along the dimension of human likeness at which category membership as human or non-human is highly ambiguous (illustration adapted from MacDorman, 2005).

by recent progress in robotics and computer graphics technologies in the realistic simulation of aspects of human appearance and behavior and, therefore, by interest in understanding the impact of enhanced anthropomorphic realism on affective experience (e.g., Ho and MacDorman, 2010; Yamada et al., 2013). But empirical support for the predicted uncanny effect has been inconsistent (e.g., Hanson, 2006; MacDorman, 2006; Tinwell and Grimshaw, 2009; MacDorman et al., 2013). This has led to the query as to how research should now best proceed (Zlotowski et al., 2013). To move beyond this seeming impasse, some of the reasons for inconsistency in findings and new avenues of approach can be considered.

One reason for inconsistency stems from the UVH' ambiguous definition of the concept *shinwakan*. Mori coined this Japanese neologism to describe in simple terms the positive and negative character of affective experience in response to variously humanlike objects (**Figure 1**); the UVH defines these objects along a *dimension of human likeness* (*DHL*). Though relating to affective experience, the poor specification of this concept has resulted in much debate and various renderings of its meaning, with the investigation of the UVH and shinwakan in terms of constructs such as pleasantness, comfort level, eeriness, familiarity (i.e., feelings of familiarity vs. strangeness), likability, and empathy (e.g., MacDorman and Ishiguro, 2006; Bartneck et al., 2007; Seyama and Nagayama, 2007, 2009; Green et al., 2008; MacDorman et al., 2009a, 2013; Tinwell and Grimshaw, 2009; Dill et al., 2012; Burleigh et al., 2013). Adding to, or reflecting, the UVH' conceptual ambiguity, Mori (2012) recently re-termed the positive and negative character of affective experience as affinity.

An alternative approach to investigating affective experience of human and humanlike entities is to consider affective experience in terms of its psychologically well-validated components. Mori (1970) uses illustrative instances of feelings and affective evaluations to describe his understanding of shinwakan. Based on these instances, shinwakan could well be considered in terms of the constructs *valence* and *arousal* because these are intrinsic to all of his examples (cf. Davitz, 1969). These are also intrinsic to the affective dimensions (e.g., likeability, pleasantness, feelings of familiarity) typically used to examine the UVH and to all of the early theoretical accounts of uncanny experience (e.g., MacDorman, 2005), and they appear to be relevant for the definition of uncanny feelings in terms of specific emotions (cf. Russell, 1980; Ho et al., 2008). Valence refers to the pleasant-to-unpleasant quality (hedonic tone) and arousal to the low-to-high degree of excitement of affective experience (Barrett and Russell, 1999; Kensinger and Corkin, 2004). Given that valence and arousal form the primary orthogonal dimensions of affective experience (Schlossberg, 1954; Russell, 1980, 2003; Yik et al., 1999; Posner et al., 2005), that emotional states can be defined in the emotional space determined by these two dimensions (Bradley and Lang, 1994; Barrett and Russell, 1999), and that the principal variance in the meaning of emotional states can be explained by valence and arousal (Osgood et al., 1957; Mehrabian and Russell, 1974; Smith and Ellsworth, 1985), it is likely that indicators of valence and arousal will provide a sound basis for examining affective experience of variously humanlike entities along the DHL and the notion of the uncanny effect.

A second reason for inconsistent findings is likely to relate to the more or less exclusive reliance in uncanny-related research on *ad hoc* developed self-rating scales to assess affective experience (e.g., Hanson, 2006; MacDorman, 2006; Green et al., 2008; MacDorman et al., 2009a,b; Seyama and Nagayama, 2009; Looser and Wheatley, 2010; Tinwell et al., 2010). *Ad hoc* self-rating scales have found favor in uncanny research because they are inexpensive and easy to administer (see, e.g., Ho et al., 2008). But the psychometric validity and reliability of these measures as indicators of *shinwakan,* or of the affective experience that the notion of *shinwakan* is thought to capture, has not been demonstrated (but for steps toward construct validation, see Ho and MacDorman, 2010). This makes the interpretation and synthesis of the research findings to date difficult.

As an alternative to *ad hoc* scales, and in keeping with the foregoing considerations on arousal and valence, well-validated measures such as the *Self-Assessment Manikin* (*SAM,* e.g., Lang, 1985; Lang et al., 1997; Bradley and Lang, 2007) could be used. The SAM is a pictorial assessment technique for self-rated measurement of valence and arousal. As a non-verbal and largely culture free measure (Morris et al., 1993; Bradley et al., 1994), it would be useful – given the preceding concerns about the ambiguous concept shinwakan – for application in uncanny-related research. Though a valuable source of information, self-ratings of affective experience (e.g., of pleasantness, comfort level, eeriness, familiarity, likability, valence, or arousal) effectively

place the focus of investigation on conscious feelings of emotional state (Barrett, 1996), that is, on the explicit cognitive evaluation of the affect-related properties of the stimuli and of the psychophysiological reactions that these stimuli elicit. It is, however, conceivable that the proposed uncanny effect might also manifest itself in affect-related reactions that escape conscious detection and evaluation (for affect-related neural processes during passive category processing of faces along the DHL, see Cheetham et al., 2011). Self-report measures can be augmented therefore by direct measures of psychophysiological reactions. For the present study, we considered the use of *facial electromyography* (*EMG*) and *electroencephalography* (*EEG*) to indicate rapid and subliminal changes in emotional state (Wu et al., 2012). Affective experience is associated with psychophysiological indices that correlate differentially with valence and arousal (Cacioppo et al., 1986; Bradley and Vrana, 1993; Lang et al., 1993, 1997; Schupp et al., 2000; Amrhein et al., 2004; Foti and Hajcak, 2009): the EMG measures of the corrugator supercilii muscle and the zygomaticus major muscle are sensitive to negative and positive valence, respectively, and the EEG-based measure of the *late positive potential* (*LPP*) to arousal (Cacioppo et al., 1986; Lang et al., 1993, 1997; Cuthbert et al., 2000; Schupp et al., 2000; Dolcos and Cabeza, 2002; Amrhein et al., 2004; Hajcak and Nieuwenhuis, 2006; Foti and Hajcak, 2009; Hajcak et al., 2009; Weinberg et al., 2012; Wu et al., 2012).

A further reason for inconsistency in findings has been suggested to relate to the conceptualization and operationalization of the DHL (Cheetham et al., 2011). Mori's illustration of the UVH uses a single human exemplar to represent the human category (Mori, 1970), as reflected in some empirical and theoretical work (e.g., Ramey, 2005; Tinwell and Grimshaw, 2009). But this conceptualization effectively assumes that there is no variation in physical or psychological similarity space within the human category of the DHL. Examination of perceptual discriminative and category processing along the DHL, operationalized using morph continua to represent a linear dimension of physical similarity space spanning between human and non-human category exemplars, shows that this assumption is incorrect (Cheetham et al., 2011, 2014). The use of linear morph continua to represent the DHL is not new (e.g., Seyama and Nagayama, 2007), but poor control of continua might have contributed to inconsistency in findings. Morph continua have been subject to various experimental confounds that are likely to have systematically biased subjective experience of objects along the DHL (for a critical discussion, see Cheetham and Jancke, 2013). These confounds range from the use of different juxtaposed morph continua to represent the DHL, thus generating perceptual discontinuities along the morph continua (Hanson et al., 2005; MacDorman and Ishiguro, 2006) to morphing noise (e.g., disparities in the alignment of facial features between successive morphs of a continuum), as indicated in a recent study of the effects on subjective experience of category ambiguity (Yamada et al., 2013). Critically, morphing noise is likely to alter the cognitive representation of the human–nonhuman category structure of the DHL and may itself influence subjective experience (Cheetham and Jancke, 2013).

Understanding the human–nonhuman category structure of the DHL provides an approach to testing the ideas underlying the vaguely formulated UVH (Cheetham et al., 2011, 2014; Burleigh et al., 2013; Yamada et al., 2013; Burleigh and Schoenherr, 2015; Ferrey et al., 2015). The UVH makes no explicit reference to perceptual and category processing or to the large body of pertinent literature. But, based on this understanding, Mori's ideas can be tested by augmenting the UVH with the assumption that the predicted state of negatively valenced affect is most likely to occur at the point of realism along the DHL at which attribution of stimuli to the human or non-human category is subject to greatest categorization ambiguity (i.e., the "valley" in **Figure 1**). A similar approach has been applied to examining the relationship between the predicted state of negatively valenced affect and the ability to perceptually discriminate between (rather than categorize) morphs along continua representing the DHL (Cheetham et al., 2014). Cheetham et al. (2014) reported a number of effects that, however, provided no support for Mori's ideas.

The aim of this study was to examine affective experience in response to the presentation of highly similar human and humanlike facial stimuli along the DHL. In view of the uncertainty surrounding the conceptual definition and translation of shinwakan, we focused in the main experiment (i.e., the first of two experiments) on the examination of valence and arousal as two of the primary properties of affective experience, using non-verbal measures. The facial *EMG* measure of the corrugator supercilii muscle and the LPP were used as psychophysiological indices of valence and arousal, respectively. The SAM was used to assess self-reported valence and arousal. The DHL was represented using morphs drawn from carefully controlled continua generated from avatar and human faces. These continua were previously tested to ensure that the cognitive representation of the category structure (i.e., morph location of categorically highly unambiguous avatar and human face exemplars and of categorically most ambiguous faces) was consistent across continua.

To test the main prediction of the UVH, we assumed that the face morph associated with greatest category ambiguity along the DHL would evoke greater negatively valenced and more arousing affective experience compared with that evoked by categorically unambiguous avatar and human morphs. In the second experiment, we examined the relationship between the DHL and shinwakan in order to provide, using our stimuli, a general reference of comparison with previous studies that have focused on shinwakan using *ad hoc* scales. For this, an *ad hoc* self-rating scale of shinwakan based on the bi-polar dimension of *familiarity* (i.e., feelings of familiarity vs. strangeness) was used. Familiarity was selected because this rendering of shinwakan has been frequently investigated and because it closely captures the essence of Mori's description of the uncanny. In keeping with the preceding considerations on category structure, and assuming for the purpose of experimentation that Mori's conjectures are correct, self-rated experience of familiarity was expected to show that feelings of strangeness would be greater along the DHL for categorically ambiguous morphs compared with categorically unambiguous avatar and human morphs.

Materials and Methods

Participants

Healthy male and female adults with no record of neurological or psychiatric illness and no current medication use volunteered for one of the two studies. Participants in the first experiment, conducted in Wurzburg, were students of the University of Wurzburg and those of the second experiment, conducted in Zurich, were students of the University of Zurich. All participants were native or fluent speakers of Standard German, consistently right-handed (Annett, 1970), and had no previous experience designing or modifying computer-generated characters in, for example, virtual reality-based role-playing games, second life, or virtual reality environments, or experience using such environments (e.g., for psychotherapy, rehabilitation, training, e-commerce, or virtual reality-based research). Written informed consent was obtained before participation according to the guidelines of the Declaration of Helsinki. Each volunteer received 20 Swiss Francs or the equivalent in Euros for participation. The study and all procedures and consent forms were approved by the Ethics Committee of the Universities of Wurzburg and Zurich.

Materials and Stimuli

Twenty linear morph continua were generated, using FantaMorph software (Version 5.3.5, Abrosoft[1]), from 20 different pairs of color images of avatar and natural human faces. Each pair represented the two endpoints of a morph continuum, the continua being used to represent the DHL (for an example of stimuli used in this study, see Figure and Supplementary Figure S1B). Each continuum comprised 13 different morphed images, labeled M0 (avatar endpoint) to M12 (human endpoint), with each morph position representing an equally spaced-point along its respective continuum at increments of 8.33%. All faces were unknown and male, showing full face, frontal view, neutral expression, direct gaze, and no salient features such as facial hair and jewelery. The modeling suite Poser 7 (Smith Micro Software[2]) was used to generate and model in detail the facial geometry and texture (e.g., age, configural cues, skin tone) of the avatar faces to closely match the corresponding human face of the respective continua. The images were then edited in Adobe Photoshop CS3 to mask external features with an elliptic form and black background (96 dpi and 560 × 650 pixels), to ensure final alignment of avatars and human facial features, and to match contrast levels and overall brightness of each pair of parent faces before morphing.

The 20 continua (260 stimuli in total) were used to ensure a sufficient number of trials for signal averaging across continua in order to enhance the signal-to-noise ratio for the LPP and EMG measures. The final choice of continua was based on three pilot studies ($N = 82$). Two of these pilot studies used a *two-alternative forced choice classification task* to verify the consistency of the category structure of the DHL across the continua. This task required that the participants identify the presented stimulus as either an avatar or human as quickly and accurately as possible after stimulus onset by pressing one of two response keys. These pilots showed that faces at morph positions M0, M1, M2, and M3 were highly unambiguously assigned to the avatar category, faces at M9, M10, M11, and M12 were highly unambiguously assigned to the human category, and that M6 was most closely associated with greatest ambiguity in categorization judgments (for details and Supplementary Figure S1A, see Supplemental Information 1). Another pilot study ($N = 18$) was used to judge the facial attractiveness of avatar and human endpoints of all continua before these were morphed. A dependent sample t-test showed that there was no significant difference in attractiveness ratings between the avatar ($M = 2.76$, SD = 0.44) and human parent images ($M = 2.87$, SD = 0.52), $t_{17} = -1.076$, $p = 0.297$.

[1]http://www.abrosoft.com

[2]http://www.smithmicro.com

Experiment 1: Psychophysiological Recordings and Ratings of Valence and Arousal

Participants

Of $N = 30$ participants, three were excluded before data analyses because of excessive impedances during facial EMG and EEG acquisition, leaving $N = 27$ participants aged between 18 and 36 years (15 female; $M = 23$ years; SD = 4.02).

Materials and Procedure

All participants were tested individually. Participants were seated in a small, sound-attenuated, dimly lit, shielded cabin, and electrodes for EMG and EEG acquisition were attached. Participants then provided demographic information and completed the *State Trait Anxiety Inventory* (*STAI*; Spielberger et al., 1970; German version by Laux et al., 1981). A 1 min resting baseline was performed at the beginning of the experiment to facilitate laboratory adaptation. Each participant received written instructions presented on the screen before commencement of each of three tasks, which were always performed in the same order (in keeping with standardized procedure, e.g., Amrhein et al., 2004).

Three tasks were conducted. Each task presented the same 260 stimuli (550 × 650 pixels) at a viewing distance of 62 cm and subtended a visual angle of 11° × 14°; this is approximately equivalent to viewing a real face from a normal distance during conversation of 90–100 cm (Hall, 1991; Henderson et al., 2005). The stimuli were always presented individually and in random order, with the constraint that no stimuli from within the same continuum or from corresponding morph positions of the different continua were shown in sequence.

In Task 1, participants viewed the stimuli for the duration of each stimulus' presentation, without any further task requirement. Each trial began with a stimulus that timed out at 750 ms. followed by the *inter-trial interval* (*ITI*). The ITI varied randomly (between 4,000 and 5,000 ms), showing a black screen with a white fixation cross. EMG and EEG psychophysiological measures were recorded concomitantly. In Task 2, the stimuli were presented to the participants using a computerized version

of the SAM. This required that participants press an appropriate key to indicate their subjective ratings of valence and arousal for each stimulus; the SAM rating scales range from 1 to 9 (i.e., very positive to very negative for valence and very high to very low for arousal). The button press was followed by the ITI (as in Task 1). A practice pre-test of five trials using stimuli from continua not included in the main test was performed to ensure correct use of the SAM rating scales. In Task 3, a *two-alternative forced choice classification task* was conducted to verify the location of the morph associated with greatest categorization ambiguity and the profile of avatar and human category decisions for the other morphs; please note that this task is the same as the two-alternative forced choice classification task used in the pilot studies. This task required that participants press an appropriate key to indicate their category judgments. The button press was followed by the ITI (as in Task 1), with time out at 750 ms. A practice pre-test of five trials using stimuli from continua not included in the main test was performed to ensure correct use of the response buttons and comprehension of the category label 'avatar.'

Psychophysiological Data Recording and Reduction

Continuous psychophysiological recording was performed in Task 1. EMG acquisition entailed bipolar placement of Ag/AgCl electrodes with surface diameter of 7 mm over the left M. corrugator supercilii (Fridlund and Cacioppo, 1986). Participants were told that skin conductance would be recorded (Dimberg et al., 2000; Weyers et al., 2006, 2009; Likowski et al., 2012). The EMG raw signal was measured with a V-Amp 16 amplifier (Brain Products Inc., Gilching, Germany) and stored with a sampling frequency of 1000 Hz. Raw data were then rectified and filtered offline with a 30 Hz low pass and 500 Hz high pass cut-off filter, a 50 Hz notch filter, and integrated with a 125 ms time constant. The EMG difference scores were computed on the basis of the mean change in activity after stimulus onset from 500 ms baseline before stimulus onset. Trials with EMG activity exceeding 8 μV during the 500 ms baseline and above 30 μV during stimulus presentation were excluded (less than 5%). For statistical analyses, the data of each participant were collapsed over the 20 trials of each of the 13 corresponding morph positions of the 20 continua and averaged over the 100 ms intervals post-stimulus onset (Weyers et al., 2006; van Boxtel, 2010).

For the LPP, EEG was recorded using Ag/AgCl electrodes placed at mid-line sites according to the international 10–20 system (i.e., sites FCz, Cz, CPz, Pz, C1, C2, CP1, CP2) at a sampling rate of 1,000 Hz and referenced to Cz during data recording and replaced by the mean of mastoids during off-line data analysis. Electrodes were mounted on an Easycap (EasyCap, Hersching, Germany). Raw data were processed and analyzed using the computer Brain Vision Analyzer software (Version 2.0, Brain Products Inc.). The continuous EEG data were subjected to band-pass between 0.01 and 20 Hz filter off-line. Trials with EEG activity exceeding a transition threshold of 50 μVolt (sample to sample) or amplitude of 300 μV were excluded from further analysis. EEG data was corrected for blinks and eye movement artifacts (Gratton et al., 1983). Data for the LPP was extracted for stimulus synchronized epochs from 100 ms baseline preceding stimulus onset till 750 ms post-stimulus onset. The data was then baseline corrected (i.e., the 100 ms before stimulus onset), and then averaged for each participant and each of the 13 corresponding morph positions of the 20 continua. The LPPs were determined on the basis of mean amplitude calculated over time windows on the basis of the literature (e.g., Schupp et al., 2000). In particular, LPPs were scored as mean activity between 300 and 750 ms after stimulus onset over the midline electrode sites (CPz, CP1, and CP2; Schupp et al., 2004; Hajcak et al., 2007, 2010; Foti and Hajcak, 2008).

All data analyses in this and the following experiment were performed using SPSS version 18.0 (SPSS, Inc., Chicago, IL, USA).

Results

STAI Questionnaire

The average score of the STAI state scale was $M = 36.07$ (SD = 7.96, range = 22–55) and that of the STAI trait scale was $M = 36.42$ (SD = 10.19, range = 20–60); the STAI scales measure anxiety, ranging from 20 (not at all anxious) to 80 (very anxious).

Categorization Responses

To verify the choice of categorically ambiguous and unambiguous morphs for further analyses, informal inspection of the results of the two-alternative forced choice classification task (Task 3 of Experiment 1), indicates that morph position M6 is associated with greatest ambiguity in avatar-versus-human categorization responses and that the avatar (i.e., M0, M1, M2, M3) and human faces (i.e., M9, M10, M11, M12) show a lower and upper asymptote that nears 95% (see Supplementary Figure S2 in Supplemental Information 2); this profile of category judgments is consistent with the pilot data (see Supplemental Information 1).

To characterize this profile more clearly, the mean categorization response data for M6 were compared with the aggregated mean data for the avatar faces and human faces. Greenhouse–Geisser adjustment was applied to correct the degrees of freedom for violation of the sphericity assumption as appropriate in this and subsequent analyses. A one-way *repeated measures of analysis of variance* (*RM-ANOVA*) with the factor *morph* position (three levels: M6, 'M0, M1, M2, M3' and 'M9, M10, M11, M12') was conducted on the dependent variable categorization *response* for each participant across continua. The categorization responses were entered in the analysis in terms of percentage of responses categorized as human. This analysis showed a highly significant effect for the expected differences for morph position in categorization responses, $F(2,52) = 347.42$, $p < 0.001$, with M6 approaching chance level of 50% in categorization responses ($M = 0.44$; SD = 0.17), whereas the morphs at M0, M1, M2, and M3 and at M9, M10, M11, and M12 were clearly judged to be exemplars of the avatar ($M = 0.04$; SD = 0.06) and human face categories ($M = 0.9$; SD = 0.12), respectively (see **Figure 2A**).

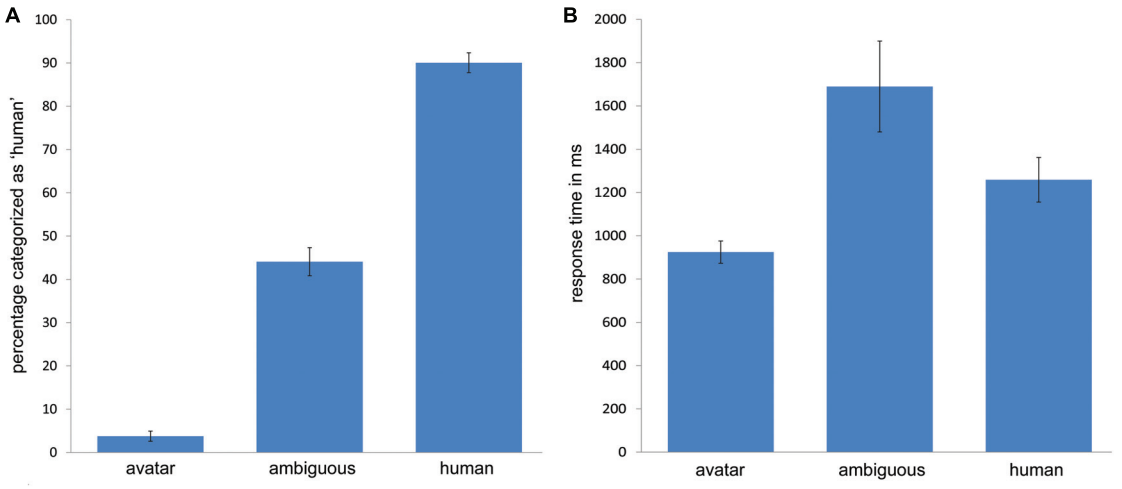


FIGURE 2 | Mean categorization responses and response times in the two-alternative forced choice categorization task. Mean categorization responses are depicted in terms of percentage of 'human' responses **(A)** for morphs representing unambiguous *avatar* and *human* faces and the most *ambiguous* face along continua representing the DHL. The depiction of corresponding mean response times for categorization responses **(B)** show longer response latencies for ambiguous faces and the shortest response latencies for avatar faces. The error bars indicate standard errors ($N = 27$).

Categorization Response Times (RTs)

The longest response latency might be expected to correspond with greatest categorization ambiguity along the DHL (Cheetham et al., 2011, 2013). To confirm this for the present data (though RT is not relevant for the Tasks 1 and 2 in Experiment 1), the same *RM-ANOVA* as in the preceding was applied using categorization *response times (RTs*; in ms) rather than categorization response as the dependent variable. Given the reported asymmetries in processing avatar and human faces along the DHL (Cheetham et al., 2011, 2013), this and the following analyses included pre-planned contrasts to compare measures for M6 and the human and avatar faces. This analysis showed a significant effect for morph position along the DHL, $F(1.11,29.06) = 17.13$, $p < 0.001$. The pre-planned contrasts showed a significant difference between RT at M6 ($M = 1689$; SD = 1091) compared with the mean average for avatar *and* human morphs ($M = 1097$; SD = 551), $F(1,26) = 20.41$, $p < 0.001$, indicating that RT is significantly longer at M6 than for other morphs. Consistent with previous RT data (Cheetham et al., 2011, 2013; Cheetham and Jancke, 2013), pre-planned contrasts showed a significant difference in RT between the avatar ($M = 924$; SD = 276) and human morphs ($M = 1258$; SD = 534), $F(1,26) = 22.58$, $p < 0.001$, such that the mean latency of categorization responses was longer for human faces (see **Figure 2B**).

Affective Experience and Categorization Ambiguity

Based on the preceding, we compared the lpp, emg, sam for valence and sam for arousal data at m6 with the data for the avatar and human category faces. Separate one-way rm-anovas with the factor *morph* position (three levels: m6, 'm0, m1, m2, m3' and 'm9, m10, m11, m12') were conducted for each of the dependent variables lpp, emg, sam ratings for valence, and sam ratings for arousal of each participant across the 20 continua.

For lpp, there was a significant effect of morph position along the DHL, $F(1.31,34.09) = 5.62$, $p = 0.016$. Pre-planned contrasts showed a significant difference between the avatar ($M = 10.92$; SD = 3.54) and human faces ($M = 9.31$; SD = 3.5), $F(1,26) = 36.31$, $p < 0.001$, such that the measure for LPP was greater for the avatar than for the human category (see **Figures 3A** and **4**). The pre-planned contrasts showed no significant difference between M6 and the avatar or the human faces, respectively. These data indicate that the LPP values increased across the three stimulus conditions (i.e., avatar faces, M6 and human faces) with increasing morph distance from the human end of the continua.

For EMG and the corrugator supercilii activity (see **Figure 3B**), there was no significant effect of morph position, $F(1.38,35.83) = 0.66$, $p = 0.47$.

There was a significant effect of morph position for the SAM arousal ratings, $F(1.51,39.38) = 3.78$, $p = 0.043$. Pre-planned contrasts showed a significant difference between the avatar morphs ($M = 3.6$; SD = 1.54) and M6 ($M = 3.38$; SD = 1.49), [$F(1,26) = 4.43$, $p = 0.045$] and between the avatar and the human morphs ($M = 3.34$; SD = 1.58), [$F(1,26) = 4.45$, $p = 0.045$], such that arousal ratings increased across the three stimulus conditions with increasing distance from the human end of the continua (see **Figure 3C**). There was no significant difference between arousal ratings for M6 and human faces.

For the SAM valence ratings, there was a highly significant effect of morph position along the DHL, $F(1.29,33.65) = 56.69$, $p < 0.001$. Pre-planned contrasts showed a significant difference between avatar faces ($M = 5.68$; SD = 0.62) and M6 ($M = 4.91$; SD = 0.51), [$F(1,26) = 59.75$, $p < 0.001$], between M6 and the human faces ($M = 4.55$; SD = 0.68), [$F(1,26) = 24.63$, $p < 0.001$], and between avatar and human faces [$F(1,26) = 36.31$, $p < 0.001$]. These data show that the valence ratings increased negatively

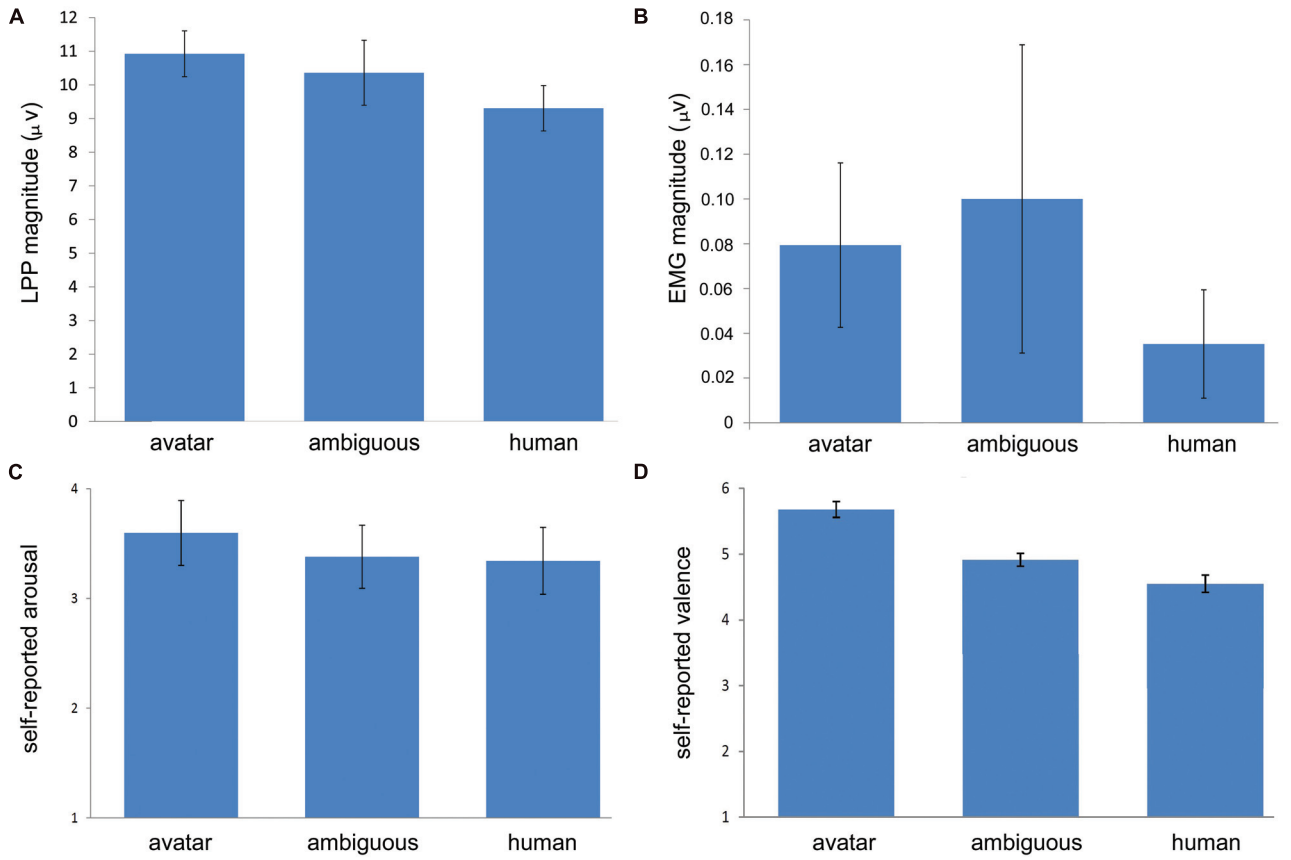


FIGURE 3 | Effect of highly unambiguously avatar and human and highly ambiguously categorized faces on arousal and valence. Overall, the figures illustrate a general increase in LPP magnitude indicating arousal **(A)**, in self-rated arousal **(C)** and in self-rated negative valence **(D)** with decreasing human likeness from the human to the avatar faces. Differences in EMG magnitude indicating negative valence **(B)** are not significant. The error bars indicate standard errors (N = 27).

across the three stimulus conditions with increasing morph distance from the human end of the continua (see **Figure 3D**).

Please note that the levels 'M0, M1, M2, M3' and 'M9, M10, M11, M12' were selected for the preceding analyses to represent the avatar and human categories. But performing the same separate one-way RM-ANOVAs using just M0, M6, and M12 as the morph factor levels produced the identical pattern of results for LPP, EMG, SAM ratings for valence, and SAM ratings for arousal.

Experiment 2: Ratings of familiarity

Participants and Procedure

A sample of N = 30 participants aged between 20 and 30 years (15 female; M = 25.64 years; SD = 2.88) were examined. The laboratory, stimulus conditions, task requirements, and instructions in this study were the same as described for Task 2 (i.e., self-ratings of valence and arousal) of Experiment 1, except that participants were required to view and rate feelings of familiarity in response to stimuli on a 5-point Likert scale by pressing the appropriate response key as quickly and accurately as possible after stimulus onset. This task permitted the analysis of RTs for familiarity judgements. The rating scale ranged from very strange (1) to very familiar (5). A practice pre-test of five trials was applied, as described for the tasks of Experiment 1. Please note that the pilot study to determine facial attractiveness of continua endpoints, described in Section "Materials and Stimuli," was the same as used for self-ratings of familiarity, except that a 5-point bipolar Likert rating scale ranging from very unattractive (1) to very attractive (5) was used.

Results

Familiarity Ratings

A one-way RM-ANOVA with the factor *morph* position (three levels: M6, 'M0, M1, M2, M3' and 'M9, M10, M11, M12') and the dependent variables *familiarity* rating of each participant across the 20 continua revealed a highly significant effect of morph position, $F(1.27,36.85) = 109.03$, $p = <0.001$. Pre-planned contrasts showed a significant difference between the avatar morphs (M = 1.9; SD = 0.72) and M6 (M = 3.01; SD = 0.54), [$F(1,29) = 134.42$, $p = <0.001$] and between M6 and the human

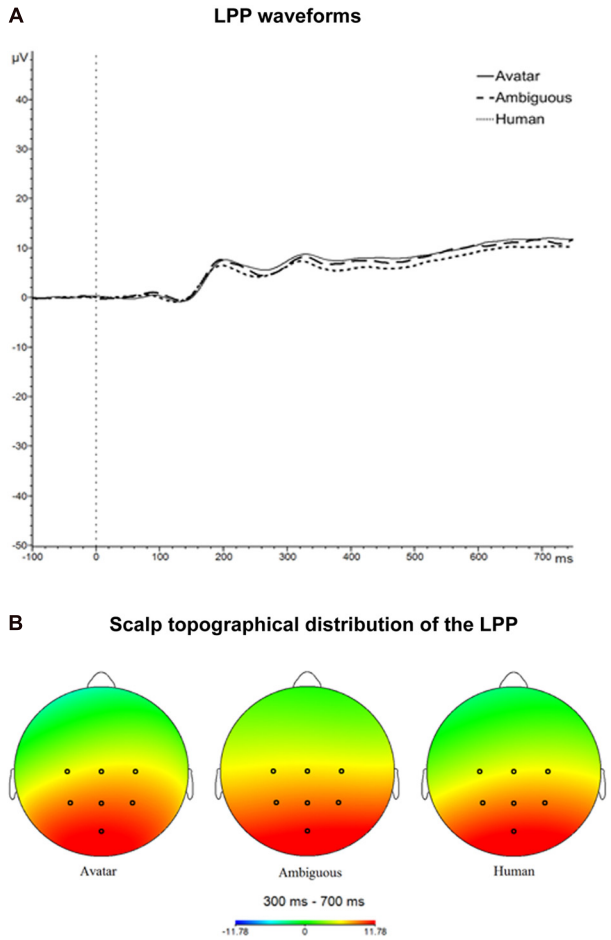


FIGURE 4 | Late positive potential (LPP) waveforms and scalp topographical distribution. The figure illustrates the LPP waveforms **(A)** and the scalp topographical distribution of the LPP **(B)** for avatar faces, ambiguous faces and human faces.

morphs (M = 3.68; SD = 0.54), $F(1,29)$ = 48.58, p = <0.001, such that familiarity ratings increased negatively across the three stimulus conditions with increasing distance from the human end of the continua (see **Figure 5A**).

RT of Familiarity Ratings

The same analysis, using instead RT for familiarity ratings also showed a highly significant effect of morph position, $F(1.27,36.85)$ = 12.71, p = <0.001 (see **Figure 5B**). Reflecting the pattern of RT for categorization in Task 3 of Experiment 1, pre-planned contrasts showed a significant difference between RT at M6 (M = 1682; SD = 527) compared with the mean average for avatar *and* human morphs (M = 1482; SD = 376), $F(1,26)$ = 20.41, p < 0.001, indicating that RT for familiarity judgments is significantly longer at M6 than for categorically unambiguous morphs. Pre-planned contrasts showed also that RT for familiarity ratings of avatars (M = 1400; SD = 453) was significantly faster than for human faces (M = 1627; SD = 437), $F(1,29)$ = 11.9, p = 0.002.

Discussion

The main prediction of the UVH is that observation of highly humanlike characters or objects that are difficult to distinguish from the human counterpart will elicit negative affect, whereas the affective experience of distinctly non-human and human characters or objects will be more positive in comparison. To test this, measures of valence and arousal (i.e., the primary orthogonal dimensions of affective experience) were used to compare the impact on affective state of categorically ambiguous faces with that of categorically unambiguous avatar and human faces; the physical morph distance of the avatar and human faces from the ambiguous faces was controlled for. To reflect Mori's idea, we assumed that any evidence in support of an uncanny-like effect (i.e., enhanced negative affective experience for ambiguous faces) during passive viewing and during explicit affect evaluation would most likely be revealed by comparing the categorically most ambiguous stimuli with the most unambiguous stimuli. But the data from the LPP, EMG, and SAM-based measures of arousal and valence converge in showing no support for the notion that category ambiguity along the DHL is specifically associated with enhanced experience of negative affect. On the contrary, the LPP and SAM-based measures indicate a general increase in arousal and negative valence across the stimulus conditions (i.e., the human category, ambiguous category, avatar category) with increasing morph distance from the human end of the continua. The *ad hoc* familiarity ratings delivered the same picture, indicating that feelings of strangeness are not specifically associated with categorically ambiguous faces and generally increase with increasing morph distance from the human end of the continua.

These findings are consistent with the profile of subjective evaluations reported in other studies that have used comparable, well-controlled morph continua and *ad hoc* measures of shinwakan, such as pleasantness (e.g., Looser and Wheatley, 2010). This profile is characterized by a general asymmetry in affective experience, with increasingly negative evaluations of morphs with decreasing human likeness. But asymmetry along the DHL is not specific to affective processing. Asymmetries have also been reported in tasks of perceptual and category processing in other uncanny-related studies (Cheetham et al., 2011, 2013, 2014). These tasks have revealed greater decision certainty and shorter RT latencies in categorization judgments for avatar compared with human faces, extraction of different perceptual details from avatar compared with human faces during perceptual decision making, differential sensitivity of affect-related brain structures (e.g., amygdala, insula) to avatar compared with human faces during passive viewing, and enhanced discrimination sensitivity to perceptual differences in visual information between faces within the avatar compared with those within the human category. The data of the present study are therefore worth considering in the context of such asymmetries in perceptual and category processing.

The RT data in the categorization task (i.e., Task 3 of Experiment 1) show that explicit categorization judgments

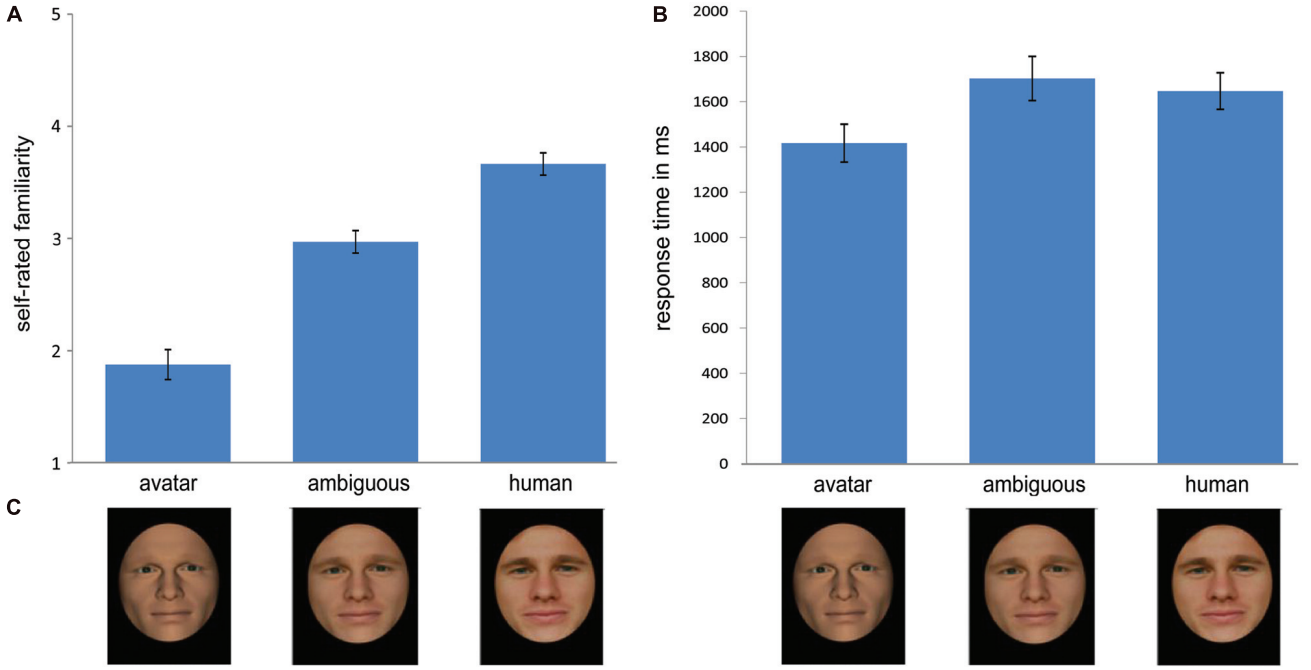


FIGURE 5 | Self-reported feelings of familiarity. (A) Illustrates a general increase in familiarity ratings with increasing human likeness from the avatar to the human faces **(A)**. Reflecting the pattern of response times for categorization response times in the two-alternative forced choice categorization task, response latencies for ambiguous faces are longer compared with the avatar and human faces. The error bars indicate standard errors ($N = 30$). **(B)** Shows examples of morphs drawn from a continuum of the kind used to represent the DHL. The morph labeled M0 corresponds with the avatar endpoint of the continuum, M6 with greatest category ambiguity, and M12 with the human endpoint of the continuum. **(C)** Shows examples of stimuli used in this study.

for avatar faces are faster than those for human faces. One interpretation of this finding is that different perceptual features are used for processing category information in novel avatar compared with "everyday" human faces, as indicated for implicit and explicit processing of perceptual and category information along the DHL in other studies (Cheetham et al., 2011, 2013, 2014). It has been suggested that this difference might relate to the use of novel and thus salient perceptual information as a readily identifiable feature of avatar faces (e.g., novel color, smoothed skin texture, or feature shape) to facilitate category processing (Cheetham et al., 2013). Assuming that novel information is easier to extract from the avatar faces compared with the corresponding category information from the human faces and that this information is preferentially used as diagnostic of avatar category membership, one might expect an *RT facilitation effect* such that RT latencies for category judgments of novel avatar faces are shorter compared with those of human faces (see Levin and Angelone, 2001). The present data are consistent with this idea (for an alternative explanation of RT facilitation, see Valentine, 1991).

While there was no requirement to categorize our stimuli during the other tasks (i.e., during passive observation, explicit evaluation of affect, and familiarity judgments), it is likely that participants did engage in implicit processing of the categories (for evidence of this during passive observation, see Cheetham et al., 2011; see also Castelli et al., 2004). If implicit processing of category did occur, a similar RT facilitation effect might be expected for our affective evaluations of avatar faces (please note that for experimental reasons, RT was only collected for familiarity judgments). This suggestion assumes that processing of category membership contributes to or influences in some way the processing of familiarity (for categorization effects and the evaluation of attractiveness, see Halberstadt and Winkielman, 2014). In fact, the data show that the latency of familiarity judgments is shorter for avatar than for human faces. The data show also that the latency of familiarity judgments is longest for the categorically most ambiguous faces. Familiarity judgments thus appear to be influenced by or interact with the processing and cognitive representation of the category structure of the continua (cf., e.g., Heekeren et al., 2008). Similar effects might apply for the psychophysiological and for the SAM-based measures of valence and arousal. The use of a different task design that allows clear interpretation of RT data of the SAM-based measures of valence and arousal might be used to examine this.

The preceding considerations hint at the possibility that there is a relationship of some kind between cognitive processing efficiency (as indicated in the preceding by RT), perceptual and category processing of the DHL, and our measures of affective experience of DHL stimuli. Further analysis of the familiarity ratings supports the idea of a relationship, revealing a highly significant correlation for avatars only, such that shorter RT for category judgments of avatars is associated

with more negative familiarity ratings (i.e., greater strangeness), $r = 0.575$, $p > 0.001$. This finding is not consistent with the recent *inhibitory-devaluation hypothesis* that has been presented as a potential explanation for the uncanny valley (Ferrey et al., 2015). Ferrey et al.'s (2015) hypothesis posits that a stimulus that is subject to competing interpretations, such as when membership of a stimulus to one or other potential category is ambiguous, will be evaluated more negatively (for further details regarding the role of inhibitory cognition in this). Based on the use of RT to indicate decision difficulty due to competing interpretations of categorically ambiguous stimuli (see also Yamada et al., 2013), our data reveal no significant relationship between RT and familiarity ratings for categorically ambiguous faces and that the only significant relationship found (i.e., for avatar faces) shows longer RT for more positive ratings. The data of both experiments in Burleigh and Schoenherr's (2015) recent study, based on ratings of eeriness, also lend no support for the inhibitory-devaluation hypothesis.

An alternative account of the uncanny effect that has attracted attention in uncanny research considers the influence of *processing fluency* on affective experience. According to the *Hedonic Fluency Model* (Winkielman et al., 2003), negative evaluations of novel or unfamiliar stimuli relate to subjective difficulty extracting diagnostic information for quick and efficient processing (see also Bradley et al., 1993; Bornstein and D'Agostino, 1994; Mendes et al., 2007). This proposal ties in well with the idea that negatively valenced experience along the DHL might be associated with category ambiguity. Yamada et al. (2013) follow the Hedonic Fluency Model and suggest on the basis of their data that lower processing fluency, as indicated in their study by categorization decision difficulty (i.e., longer RT) for categorically ambiguous stimuli of the DHL, is associated with enhanced negative judgments of likeability. In contrast, our data indicate that any change in arousal, valence and familiarity is not modulated by effects of category ambiguity and, specifically in relation to the correlative result between RT and the faces of the avatar category, do not favor the hedonic fluency account.

It might be argued that the correlative relationship between RT and familiarity ratings for avatars fits more closely with a different model of processing fluency, the *Fluency Amplification Model* (Albrecht and Carbon, 2014). Albrecht and Carbon show that stimuli with a comparatively neutral or negative valence at the outset are not liked any more under conditions of higher processing fluency than they are under conditions of lower processing fluency. They show also that higher processing fluency of negative stimuli can actually enhance negative evaluation. Our data are consistent with the possibility that higher processing fluency of avatar faces (i.e., shorter RT latencies) and enhanced negative evaluation might be related in this way. This possibility is reinforced by the data of a recent study showing that higher processing fluency, indicated by reduced difficulty in perceptual discrimination between highly similar faces along the DHL, correlates with negative affect evaluations of familiarity (i.e., enhanced feelings of strangeness; Cheetham et al., 2014); please note that this finding is entirely contrary to the effect predicted on the basis of the UVH.

The reported increase in negative valence and arousal for avatar compared with the human faces in the present study might simply relate to an innate predisposition to treat the unfamiliar with caution (Zajonc, 1998). The strength of caution diminishes as further exposure reveals that the unfamiliar is non-threatening (Lee, 2001). Correspondingly, the relatively more positive evaluations of valence for human faces might simply reflect the impact of repeated exposure to human category examplars. Previous social interaction and the often more positive affective tone of interaction with a particular in-group is thought to lead to automatic activation of more positive evaluations (Reis and Gable, 2003; Claypool et al., 2007; Garcia-Marques et al., 2010). This *mere-exposure* effect (Zajonc, 1968; Monahan et al., 2000; Zajonc, 2001) might underpin the general increase in pleasantness and liking ratings with increasing human likeness of faces in other studies (e.g., Looser and Wheatley, 2010; Cheetham et al., 2014; see Experiment 1 in Seyama and Nagayama, 2007), such that more humanlike faces (or their human-specifying perceptual features) are evaluated as more likeable (see, Moreland and Zajonc, 1982). A recent study investigated the idea that the frequency of exposure to the faces of fictive beasts modulates affective ratings of eeriness (Burleigh and Schoenherr, 2015), but the authors report only nearly significant effects. Given that these fictive faces were not manipulated in terms of human likeness, the potential impact of mere-exposure on affective ratings of highly humanlike faces is open to further consideration.

One possible consideration is that the mere-exposure effect is mediated by the history of normal social interaction and a tendency to individuate in-group but not out-group members (Ostrom et al., 1993). This *differential processing bias* might also apply when processing human and humanlike faces (see Cheetham et al., 2013). This bias means that human participants preferentially code other members of the human in-group (i.e., our human stimuli) by directing cognitive processing resources toward more in-depth processing of facial information to enable individuation (i.e., processing at the exemplar level). In contrast, the processing of out-group members (i.e., our highly humanlike avatar faces) might be biased toward facial information that enhances detection of faces at the category level; for this kind of out-group bias by other names, see the *other-race hypothesis* (Levin, 2000), *differential processing hypothesis* (Ostrom et al., 1993), and the *other-race effect* (Rhodes et al., 2009). More in-depth processing for individuation would be consistent with the longer RT latencies for categorization of our human category faces in Task 3 of Experiment 1, since longer latency suggests more time-consuming processing of finer perceptual details (Schyns and Murphy, 1994; Lamberts, 1998; Johansen and Palmeri, 2002). Assuming that longer RT of familiarity ratings for human faces also reflects more in-depth processing, the allocation of more attentional resources needed for this might be sufficient to strengthen any effects of mere-exposure on positive evaluations of faces (Huang and Hsieh, 2013). This explanation is consistent with the present data and it could be investigated further in relation to categorization

performance. It is worth noting, however, that Cheetham et al. (2014) did already test the *differential processing bias* as a potential explanation for their finding of an asymmetry along the DHL in perceptual discrimination, that study showing enhanced perceptual discrimination of avatar faces compared with human faces. The data in that study did not support this explanation, and, in terms of perceptual discrimination, lend little support to Schoenherr and Burleigh's (2015) suggestion that an out-group bias might also underpin certain social-cultural phenomena that resemble elements of the uncanny valley idea.

We applied a two-dimensional approach to examining affective experience by placing the focus in the main experiment on arousal and valence. The data show a consistent pattern in the relationship between lesser degrees of human likeness and greater arousal (i.e., in the LPP and self-rating SAM measures) and more negative valence (i.e., in the self-rating SAM measure). The combination of negative valence and increased arousal is understood as indicating negative affective experience (Lang et al., 2005). But the EMG measure of corrugator supercilii activity showed no effects along the DHL. A straightforward interpretation of this would be that human likeness along the DHL has no differential impact on the valence of actual affective state during passive viewing. But in keeping with Larsen et al. (2003), who report a strong relationship between corrugator supercilii activity and self-reported valence ratings, further analysis of the data of each participant across continua also shows a strong positive relationship between increasing corrugator activity and more negative valence ratings ($r = 0.474$, $p = 0.006$). However, this effect is specific for the human category, suggesting that the valence of actual feeling state (as indexed by the EMG measure) and the cognitive evaluation of feeling state (as indexed by SAM measure) strongly converge for human faces only. This convergence might reflect a close coupling of the representation and integration of affective and cognitive processing of actual feeling state and of the cognitive appraisal of that state in relation to external input (Schwarz and Clore, 1983, 2006) that might be acquired through repeated exposure and social interactive experience with human others. This human-specific effect might relate to the processing of facial mimicry acquired in normal social interaction with human others (Weyers et al., 2006). The comparatively weak effect for non-human faces might thus reflect an attenuated responsiveness to the neutral expression of non-human faces.

The data show a significantly greater (i.e., more positive-going) LPP for the avatar compared with the human category faces. It is possible that a larger number of trials (i.e., use of more continua) might have increased the chance of also finding a significant difference between these two face categories and the ambiguous faces. Huffmeijer et al. (2014) recommends 30 trials to capture the LPP, whereas Moran et al. (2013) demonstrate that the LPP is stable and can be quantified with as few as 12 trials. Based on previous studies (Cuthbert et al., 2000; Yen et al., 2010; Herbert et al., 2013; Moran et al., 2013), we considered 20 continua to be adequate to test Mori's ideas. Considered in the context of the general decrease in LPP amplitude with increasing human likeness of the morphs, it is likely that any such difference between the ambiguous and the unambiguous faces would still reflect this general decrease and not produce an uncanny-like effect. But further investigations could examine this possibility by using more trials per morph level. It should be noted also that while the LPP is modulated by arousing (negatively and positively valenced) stimuli (e.g., Schupp et al., 2000; Leite et al., 2012; Weinberg et al., 2012), valence can also modulate the LPP (e.g., Cuthbert et al., 2000; Delplanque et al., 2006). The valence effect is less consistently reported (Olofsson et al., 2008), but we cannot exclude the possibility that effects of both arousal and valence are reflected in the LPP data. This would not change the findings of the present study, as arousal and valence are primary dimensions of affective experience.

The use of psychophysiological indices of affect introduces additional sources of data to the investigation of Mori's ideas. Together with the SAM-based ratings, this approach delivers a more complete picture of the underlying components of affective experience. In the present study, this picture includes direct and objective measures of psychophysiological reactions to the DHL stimuli and indirect and subjective measures based on the cognitive evaluation of the stimuli and of the consciously detected psychophysiological reactions that these evoke. The appeal of psychophysiological indices is that their measurement is conducted continuously and in real time, meaning that physiological events can be detected as they unfold over time at a temporal resolution on a millisecond scale. For example, the LPP develops at around 300–400 ms after stimulus onset, peaks at around 700 ms, and lasts for up to 6 s in total (Cuthbert et al., 2000). Similarly, EMG can provide effective measurement of minuscule and rapid changes, including changes that escape detection by the naked eye (e.g., Cacioppo et al., 1986; Dimberg, 1990; Weyers et al., 2006; Gomez et al., 2009; Mauss and Robinson, 2009; van Boxtel, 2010; Likowski et al., 2011; Wu et al., 2012). In view of the poor construct definition of the affective dimension described in the UVH, that some individuals find it difficult to conceptualize and quantify their emotional experiences (Mauss and Robinson, 2009), and that (*ad hoc*) self-reports might not well capture the constructs that they are intended to measure (Ho et al., 2008), further use of psychophysiological measures might contribute to a clearer characterization of the relationship between stimuli defined along the DHL and affect.

To characterize this relationship and to enable comparison between studies, the way in which the DHL is represented is an important consideration. One approach to its representation is to use morph continua (e.g., Seyama and Nagayama, 2007; for an alternative approach, see, e.g., Ho et al., 2008; Tinwell, 2009). The use of morphing permits close examination of the relationship between affect and experimentally controlled fine-grained differences in humanlike appearance. The general increase in positive affect toward the human end of the DHL indicated by our LLP, SAM arousal, and SAM valence (and by our familiarity) measures is reflected in the data of other studies based on subjective ratings of affect and on comparable nonhuman–human morph continua (e.g., Experiment 1 in Seyama and Nagayama, 2007; Looser and Wheatley, 2010; Cheetham et al., 2014). But there are findings inconsistent with the present data. The most similar study is that of Yamada et al. (2013).

They examined the relation between explicit ratings of likeability and category ambiguity along a morph continuum generated from the face of the cartoon character Charlie Brown and a human face. While they reported negative affect specifically in relation to category ambiguity, they indicate also that morphing disparities in the alignment of facial features may have influenced subjective ratings. Inspection of their stimuli suggests that these disparities particularly affected the categorically most ambiguous morphs by creating the appearance of a facial scar across the forehead of the morphed faces. This kind of morphing artifact is likely to have had a systematic effect on subjective ratings as it is related to the morph distance from the continua endpoints (see Cheetham and Jancke, 2013).

MacDorman and Ishiguro (2006) also used non-human and human morphs to represent the DHL, reporting an uncanny valley-like negative peak in affective ratings. But their use of more than one juxtaposed continuum to represent the DHL combined with non-equivalent increments of physical change between the morphs of the DHL (see also Hanson et al., 2005) appears to have created the stimulus conditions needed to generate an uncanny-like effect in the profile of subjective responses. It should be noted that the experimental manipulation of facial features along morph continua for the explicit purpose of evoking uncanny-like effects has been applied in other experiments (see Experiments 2 and 3 in Seyama and Nagayama, 2007). Other studies have used experimentally controlled morph continua and examined affect along similar dimensions of facial human likeness. However, these have used computer-generated rather than natural human faces to represent the human end of the human likeness dimension (MacDorman et al., 2009a). Considered in terms of the UVH, the human endpoints in these studies are in effect exemplars of the non-human category. Use of computer-generated faces to represent the human faces is not unusual in face research (e.g., Todorov et al., 2009). But there are, as discussed in the preceding, differences in perceptual and category processing between natural human and computer-generated human-like faces. Importantly, perceptual and category processing of human category and similar computer-generated, non-human category faces can correlate differently with measures of affect, as shown for perceptual discrimination (Cheetham et al., 2014) and in the present study. In terms of the UVH, this makes comparison between such studies and those that use natural human faces difficult. Burleigh et al. (2013) and Burleigh and Schoenherr (2015) have also investigated the UVH using experimentally controlled morph continua, but their face morphs were not manipulated in terms of human likeness.

The present study focused on affective experience in terms of the dimensions arousal and valence. Affective experience can be conceptualized in other ways (see e.g., Davidson, 1993; Panksepp, 1998; Lindquist et al., 2012), and examination of arousal and valence is not new in emotion research (e.g., Lane et al., 1999; Robinson and Compton, 2006; Lewis et al., 2007; Demanet et al., 2011). But this two-dimensional approach is different than that taken in uncanny research to date (e.g., Ho et al., 2008; Tinwell, 2009; Ho and MacDorman, 2010). For example, Ho et al. (2008) focus on defining specific emotions, such as fear, with which to characterize uncanny experience. But it is worth noting that Ho et al. (2008) see parallels between their own findings (using robotic stimuli), arousal and valence, and Russell's *circumplex model of affect* (Russell, 1980). This model understands affective states, such as fear, as arising from neurophysiological systems that relate to arousal and valence (for a detailed review, see Posner et al., 2005). It is conceivable that measures of arousal and valence might explain a significant amount of the variance in and provide further insight into the affective constructs (e.g., likeability, feelings of familiarity, fear, disgust) typically used to investigate the uncanny effect (for overviews, see Ferguson and Bargh, 2003; Posner et al., 2005; Cunningham and Zelazo, 2007; Schwarz, 2007). We note, however, that psychophysiological measures do not replace measures of self-reported feelings, because self-reports and measures of psychophysiological reactivity and behavior are all relevant to the description of an emotional response (Mandler et al., 1961; Lang, 1989). Whether the present findings might generalize to fine-grained manipulations of the DHL based on computer-generated stimuli using different software to generate avatars (with manipulations of different perceptual information), female face stimuli, emotionally expressive faces, and dynamic stimuli (see Chaminade et al., 2007) is open to further investigation.

Acknowledgments

This work was supported by the European Union FET Integrated Project PRESENCCIA (Contract number 27731) and by the German Research Foundation (GRK 1253/1 scholarship to LW, FOR 605-PA 566/9-1, and SFB-TRR 58 project B01).

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2015.00981

References

Albrecht, S., and Carbon, C. (2014). The fluency amplification model: fluent stimuli show more intense but not evidently more positive evaluations. *Acta Psychol.* 148, 195–203. doi: 10.1016/j.actpsy.2014.02.002

Amrhein, C., Muhlberger, A., Pauli, P., and Wiedemann, G. (2004). Modulation of event-related brain potentials during affective picture processing: a complement to startle reflex and skin conductance response? *Int. J. Psychophysiol.* 54, 231–240. doi: 10.1016/j.ijpsycho.2004.05.009

Annett, M. (1970). A classification of hand preference by association analysis. *Br. J. Psychol.* 61, 303–321. doi: 10.1111/j.2044-8295.1970.tb01248.x

Barrett, L. F. (1996). Hedonic tone, perceived arousal, and item desirability: three components of affective experience. *Cogn. Emot.* 10, 47–68. doi: 10.1080/026999396380385

Barrett, L. F., and Russell, J. A. (1999). The structure of current affect: controversies and emerging consensus. *Curr. Dir. Psychol. Sci.* 8, 10–14. doi: 10.1111/1467-8721.00003

Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. (2007). "Is the uncanny valley an uncanny cliff?," in *Proceedings of the 16th IEEE, RO-MAN,* Jeju, 368–373. doi: 10.1109/roman.2007.4415111

Bornstein, R. F., and D'Agostino, P. R. (1994). The attribution and discounting of perceptual fluency: preliminary tests of a perceptual fluency/attributional model of the mere exposure effect. *Soc. Cogn.* 12, 103–128. doi: 10.1521/soco.1994.12.2.103

Bradley, M. M., Greenwald, M. K., and Hamm, A. (1994). "Affective picture processing," in *The Structure of Emotion: Psychophysiological, Cognitive, and Clinical Aspects*, eds N. Birbaumer and A. Ohman (Toronto: Hugute-Huber).

Bradley, M. M., and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25, 49–59. doi: 10.1016/0005-7916(94)90063-9

Bradley, M. M., and Lang, P. J. (2007). "The international affective picture system (IAPS) in the study of emotion and attention," in *Handbook of Emotion Elicitation and Assessment*, eds J. A. Coan and J. J. B. Allen (Oxford: Oxford University Press), 29–46.

Bradley, M. M., Lang, P. J., and Cuthbert, B. N. (1993). Emotion, novelty, and the startle reflex: habituation in humans. *Behav. Neurosci.* 107, 970–980. doi: 10.1037//0735-7044.107.6.970

Bradley, M. M., and Vrana, S. R. (1993). "The startle probe in the study of emotion and emotional disorders," in *The Structure of by Motivational Relevance Psychophysiology*, eds N. Birnbaumer and A. Oehnman (Seattle, WA: Hogrefe & Huber), 257–261.

Burleigh, T. J., and Schoenherr, J. R. (2015). A reappraisal of the uncanny valley: categorical perception or frequency-based sensitization? *Front. Psychol.* 5:1456. doi: 10.3389/fpsyg.2014.01456

Burleigh, T. J., Schoenherr, J. R., and 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. *Comput. Hum. Behav.* 29, 759–771. doi: 10.1016/j.chb.2012.11.021

Cacioppo, J. T., Petty, R. E., Losch, M. E., and Kim, H. S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *J. Pers. Soc. Psychol.* 50, 260–268. doi: 10.1037/0022-3514.50.2.260

Castelli, L., Zogmaister, C., Smith, E. R., and Arcuri, L. (2004). On the automatic evaluation of social exemplars. *J. Pers. Soc. Psychol.* 86, 373–387. doi: 10.1037/0022-3514.86.3.373

Chaminade, T., Hodgins, J., and Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters' actions. *Soc. Cogn. Affect. Neurosci.* 2, 206–216. doi: 10.1093/scan/nsm017

Cheetham, M., and Jancke, L. (2013). Perceptual and category processing of the Uncanny Valley Hypothesis' dimension of human likeness: some methodological issues. *J. Vis. Exp.* 76:e4375. doi: 10.3791/4375

Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., and Jancke, L. (2013). Category processing and the human likeness dimension of the Uncanny Valley Hypothesis: Eye-Tracking Data. *Front. Psychol.* 4:108. doi: 10.3389/fpsyg.2013.00108

Cheetham, M., Suter, P., and Jancke, L. (2011). The human likeness dimension of the "uncanny valley hypothesis": behavioral and functional MRI findings. *Front. Hum. Neurosci.* 5:126. doi: 10.3389/fnhum.2011.00126

Cheetham, M., Suter, P., and Jancke, L. (2014). Perceptual discrimination difficulty and familiarity in the Uncanny Valley: More like a 'Happy Valley'. *Front. Psychol.* 5:1219. doi: 10.3389/fpsyg.2014.01219

Claypool, H., Hugenberg, K., Housley, M., and Mackie, D. M. (2007). Familiar eyes are smiling: on the role of familiarity in the perception of facial affect. *Eur. J. Soc. Psychol.* 37, 856–866. doi: 10.1002/ejsp.422

Cunningham, W. A., and Zelazo, P. D. (2007). Attitudes and evaluations: a social cognitive neuroscience perspective. *Trends Cogn. Sci.* 11, 97–104. doi: 10.1016/j.tics.2006.12.005

Cuthbert, B., Schupp, H., Bradley, M., Birbaumer, N., and Lang, P. (2000). Brain potentials in affective picture processing: covariation with autonomic arousal and affective report. *Biol. Psychol.* 52, 95–111. doi: 10.1016/S0301-0511(99)00044-7

Davidson, R. J. (1993). "The neuropsychology of emotion and affective style," in *Handbook of Emotions,* eds M. Lewis and J. M. Haviland (New York: Guilford), 143–154.

Davitz, J. R. (1969). *The language of emotion*. New York: Academic Press.

Delplanque, S., Silvert, L., Hot, P., Rigoulot, S., and Sequeira, H. (2006). Arousal and valence effects on event-related P3a and P3b during emotional categorization. *Int. J. Psychophysiol.* 60, 315–322. doi: 10.1016/j.ijpsycho.2005.06.006

Demanet, J., Liefooghe, B., and Verbruggen, F. (2011). Valence, arousal, and cognitive control: a voluntary task-switching study. *Front. Psychol.* 2:336. doi: 10.3389/fpsyg.2011.00336

Dill, V., Flach, L., Hocevar, R., Lykawka, C., Musse, S., and Pinho, M. (2012). "Evaluation of the uncanny valley in CG characters," in *Intelligent Virtual Agents Lecture Notes in Computer Science*, eds N. Yukiko, M. Neff, A. Paiva, and M. Walker (Berlin: Springer), 511–513.

Dimberg, U. (1990). Facial electromyographic reactions and autonomic activity to auditory stimuli. *Biol. Psychol.* 31, 137–147. doi: 10.1016/0301-0511(90)90013-M

Dimberg, U., Thunberg, M., and Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychol. Sci.* 11, 86–89. doi: 10.1111/1467-9280.00221

Dolcos, E., and Cabeza, R. (2002). Event-related potentials of emotional memory: encoding pleasant, unpleasant, and neutral pictures. *Cogn. Affect. Behav. Neurosci.* 2, 252–263. doi: 10.3758/CABN.2.3.252

Ferguson, M. J., and Bargh, J. A. (2003). "The constructive nature of automatic evaluation," in *The Psychology of Evaluation: Affective Processes in Cognition and Emotion,* eds J. Musch and K. C. Klauer (Mahwah, NJ: Lawrence Erlbaum Associates, Inc.), 169–188.

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

Foti, D., and Hajcak, G. (2008). Deconstructing reappraisal: descriptions preceding arousing pictures modulate the subsequent neural response. *J. Cogn. Neurosci.* 20, 977–988. doi: 10.1162/jocn.2008.20066

Foti, D., and Hajcak, G. (2009). Depression and reduced sensitivity to non-rewards versus rewards: evidence from event-related potentials. *Biol. Psychol.* 81, 1–8. doi: 10.1016/j.biopsycho.2008.12.004

Fridlund, A. J., and Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology* 23, 567–589. doi: 10.1111/j.1469-8986.1986.tb00676.x

Garcia-Marques, T., Mackie, D. M., Claypool, H. M., and Garcia-Marques, L. (2010). Is it familiar or positive? Mutual facilitation of response latencies. *Soc. Cogn.* 28, 205–218. doi: 10.1521/soco.2010.28.2.205

Gomez, P., Zimmermann, P. G., Guttormsen Schär, S., and Danuser, B. (2009). Valence lasts longer than arousal: persistence of induced moods as assessed by psychophysiological measures. *J. Psychophysiol.* 23, 7–17. doi: 10.1027/0269-8803.23.1.7

Gratton, G., Coles, M. G., and Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalogr. Clin. Neurophysiol.* 55, 468–484. doi: 10.1016/0013-4694(83)90135-9

Green, R. D., MacDorman, K. F., Ho, C.-C., and Vasudevan, S. K. (2008). Sensitivity to the proportions of faces that vary in human likeness. *Comput. Hum. Behav.* 24, 2456–2474. doi: 10.1016/j.chb.2008.02.019

Hajcak, G., Dunning, J. P., and Foti, D. (2007). Neural response to emotional pictures is unaffected by concurrent task difficulty: an event-related potential study. *Behav. Neurosci.* 121, 1156–1162. doi: 10.1037/0735-7044.121.6.1156

Hajcak, G., Dunning, J. P., and Foti, D. (2009). Motivated and controlled attention to emotion: time-course of the late positive potential. *Clin. Neurophysiol.* 120, 505–510. doi: 10.1016/j.clinph.2008.11.028

Hajcak, G., MacNamara, A., and Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: an integrative review. *Dev. Neuropsychol.* 35, 129–155. doi: 10.1080/87565640903526504

Hajcak, G., and Nieuwenhuis, S. (2006). Reappraisal modulates the electrocortical response to unpleasant pictures. *Cogn. Affect. Behav. Neurosci.* 6, 291–297. doi: 10.3758/CABN.6.4.291

Halberstadt, J., and Winkielman, P. (2014). Easy on the eyes, or hard to categorize: classification difficulty decreases the appeal of facial blends. *J. Exp. Soc. Psychol.* 50, 175–183. doi: 10.1016/j.jesp.2013.08.004

Hall, G. (1991). *Perceptual and Associative Learning*. Oxford: Clarendon Press. doi: 10.1093/acprof:oso/9780198521822.001.0001

Hanson, D. (2006). Exploring the aesthetic range for humanoid robots. *Paper Presented at the Cogscience, Workshop: Toward Social Mechanisms of Android Science*. Vancouver, GD.

Hanson, D., Olney, A., Prilliman, S., Mathews, E., Zielke, M., Hammons, D., et al. (2005). "Upending the uncanny valley," in *Proceedings of the 20th National Conference on Artificial Intelligence (AAAI'05)*, Vol. 4, ed. A. Cohn (Pittsburgh, PA: AAAI Press), 1728–1729.

Heekeren, H. R., Marrett, S., and Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nat. Rev. Neurosci.* 9, 467–479. doi: 10.1038/nrn2374

Henderson, J., Williams, C., and Falk, R. (2005). Eye movements are functional during face learning. *Mem. Cogn.* 33, 98–106. doi: 10.3758/BF03195300

Herbert, C., Sfärlea, A., and Blumenthal, T. (2013). Your emotion or mine: labeling feelings alters emotional face perception - an ERP study on automatic and intentional affect labeling. *Front. Hum. Neurosci.* 7:378. doi: 10.3389/fnhum.2013.00378

Ho, C.-C., and MacDorman, K. (2010). Revisiting the uncanny valley theory: developing and validating an alternative to the Godspeed indices. *Comput. Hum. Behav.* 26, 1508–1518. doi: 10.1016/j.chb.2010.05.015

Ho, C. C., MacDorman, K. F., and Pramono, Z. A. D. (2008). "Human emotion and the uncanny valley: a GLM, MDS, and isomap analysis of robot video ratings," in *Proceedings of the ACM/IEEE International Conference on Human–Robot Interaction*, Amsterdam, 169–176. doi: 10.1145/1349822.1349845

Huang, Y. F., and Hsieh, P. J. (2013).The mere exposure effect is modulated by selective attention but not visual awareness. *Vision Res.* 91, 56–61. doi: 10.1016/j.visres.2013.07.017

Huffmeijer, R., Bakermans-Kranenburg, M. J., Alink, L. R., and van Ijzendoorn, M. H. (2014). Reliability of event-related potentials: the influence of number of trials and electrodes. *Physiol. Behav.* 10, 13–22. doi: 10.1016/j.physbeh.2014.03.008

Johansen, M. K., and Palmeri, T. J. (2002). Are there representational shifts during category learning? *Cogn. Psychol.* 45, 482–553. doi: 10.1016/S0010-0285(02)00505-4

Kensinger, E. A., and Corkin, S. (2004). Two routes to emotional memory: distinct neural processes for valence and arousal. *Proc. Natl. Acad. Sci. U.S.A.* 101, 3310–3115. doi: 10.1073/pnas.0306408101

Lamberts, K. (1998). The time course of categorization. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 695–711. doi: 10.1037/0278-7393.24.3.695

Lane, R. D., Chua, P. M., and Dolan, R. J. (1999). Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. *Neuropsychologia* 37, 989–997. doi: 10.1016/S0028-3932(99)00017-2

Lang, P. J. (1985). *The Cognitive Psychophysiology of Emotion: Anxiety and the Anxiety Disorders*. Hillsdale, NJ: Lawrence Erlbaum.

Lang, P. J. (1989). "What are the data of emotion?," in *Cognitive Perspectives on Emotion and Motivation*, eds V. Hamilton, G. H. Bower, and N. Frijda (Boston, MA: Martinus Nijhoff), 173–191.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1997). "Motivated attention: Affect, activation, and action," in *Attention and Orienting: Sensory and Motivational Processes*, eds P. J. Lang, R. F. Simons, and M. T. Balaban (Mahwah, NJ: Erlbaum), 97–135.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (2005). *International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual*. Technical Report A–8, University of Florida, Gainesville, FL.

Lang, P. J., Greenwald, M. K., Bradley, M. M., and Hamm, A. O. (1993). Looking at pictures: affective, facial, visceral, and behavioral reactions. *Psychophysiology* 30, 261–273. doi: 10.1111/j.1469-8986.1993.tb03352.x

Larsen, J. T., Norris, C. J., and Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology* 40, 776–785. doi: 10.1111/1469-8986.00078

Laux, L., Glanzmann, P., Schaffner, P., and Spielberger, C. D. (1981). *Das State-Trait-Angstinventar (Testmappe mit Handanweisung, Fragebogen STAI-G Form X 1 und Fragebogen STAI-G Form X 2)*. Weinheim: Beltz.

Lee, A. Y. (2001). The mere exposure effect: an uncertainty reduction explanation revisited. *Pers. Soc. Psychol. Bull.* 27, 1255–1266. doi: 10.1177/01461672012710002

Leite, J., Carvalho, S., Galdo-Alvarez, S., Alves, J., Sampaio, A., and Gonçalves, O. F. (2012). Affective picture modulation: valence, arousal, attention allocation and motivational significance. *Int. J. Psychophysiol.* 83, 375–381. doi: 10.1016/j.ijpsycho.2011.12.005

Levin, D. T. (2000). Race as a visual feature: using visual search and perceptual discrimination tasks to understand face categories and the cross race recognition deficit. *J. Exp. Psychol. Gen.* 129, 559–574. doi: 10.1037/0096-3445.129.4.559

Levin, D. T., and Angelone, B. L. (2001). Visual search for a socially defined feature: what causes the search asymmetry favoring cross-race faces? *Percept. Psychophys.* 63, 423–435.

Lewis, P. A., Critchley, H. D., Rotshtein, P., and Dolan, R. J. (2007). Neural correlates of processing valence and arousal in affective words. *Cereb. Cortex* 17, 742–748. doi: 10.1093/cercor/bhk024

Likowski, K. U., Mühlberger, A., Gerdes, A. B., Wieser, M. J., Pauli, P., and Weyers, P. (2012). Facial mimicry and the mirror neuron system: simultaneous acquisition of facial electromyography and functional magnetic resonance imaging. *Front. Hum. Neurosci.* 6:214. doi: 10.3389/fnhum.2012.00214

Likowski, K. U., Mühlberger, A., Seibt, B., Pauli, P., and Weyers, P. (2011). Processes underlying congruent and incongruent facial reactions to emotional facial expressions in a competitive vs. cooperative setting. *Emotion* 11, 457–467. doi: 10.1037/a0023162

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., and Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. *Behav. Brain Sci.* 35, 121–143. doi: 10.1017/S0140525X11000446

Looser, C. E., and Wheatley, T. (2010). The tipping point of animacy: how, when, and where we perceive life in a face. *Psychological. Sci.* 21, 1854–1862. doi: 10.1177/0956797610388044

MacDorman, K. (2005). "Androids as an experimental apparatus: why is there an uncanny valley and can we exploit it?" in *Proceedings of the Toward Social Mechanisms of Android Science: A CogSci 2005 Workshop,* Stresa, 106–118.

MacDorman, K. F. (2006). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: an exploration of the uncanny valley. *KF MacDorman. ICCS/CogSci-* 2006.

MacDorman, K., Green, R., Ho, C.-C., and Koch, C. (2009a). Too real for comfort? Uncanny responses to computer generated faces. *Comput. Hum. Behav.* 25, 695–710. doi: 10.1016/j.chb.2008.12.026

MacDorman, K. F., Vasudevan, S. K., and Ho, C.-C. (2009b). Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *AI Soc.* 23, 485–510. doi: 10.1007/s00146-008-0181-2

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

MacDorman, K. F., Srinivas, P., and Patel, H. (2013). The uncanny valley does not interfere with level 1 visual perspective taking. *Comput. Hum. Behav.* 29, 1671–1685. doi: 10.1016/j.chb.2013.01.051

Mandler, G., Mandler, J. M., Kremen, I., and Sholiton, R. (1961). The response to threat: relations among verbal and physiological indices. *Psychol. Monogr.* 75, 22. doi: 10.1037/h0093803

Mauss, I. B., and Robinson, M. D. (2009). Measures of emotion: a review. *Cogn. Emot.* 23, 209–237. doi: 10.1080/02699930802204677

Mehrabian, A., and Russell, J. (1974). *An Approach to Environmental Psychology*. Cambridge, MA: Institute of Technology Press.

Mendes, W. B., Blascovich, J., Hunter, S. B., Lickel, B., and Jost, J. T. (2007). Threatened by the unexpected: physiological responses during social interactions with expectancy-violating partners. *J. Pers. Soc. Psychol.* 92, 698–716. doi: 10.1037/0022-3514.92.4.698

Monahan, J. L., Murphy, S. T., and Zajonc, R. B. (2000). Subliminal mere exposure: Specific, general, and diffuse effects. *Psychol. Sci.* 11, 462–466. doi: 10.1111/1467-9280.00289

Moran, T. P., Jendrusina, A. A., and Moser, J. S. (2013). The psychometric properties of the late positive potential during emotion processing and regulation. *Brain Res.* 1516, 66–67. doi: 10.1016/j.brainres.2013.04.018

Moreland, R. L., and Zajonc, R. B. (1982). Exposure effects in person perception: familiarity, similarity, and attraction. *J. Exp. Soc. Psychol.* 18, 395–415. doi: 10.1016/0022-1031(82)90062-2

Mori, M. (1970). "Bukimi no tani [The uncanny valley]. Energy, 7(4) 33-35. (Translated by Karl F. MacDorman and Takashi Minato in 2005) within

Appendix B for the paper Androids as an Experimental Apparatus: Why is there an uncanny and can we exploit it?" in *Proceedings of the CogSci-2005 Workshop: Toward Social Mechanisms of Android Science*, Italy, 106–118.

Mori, M. (2012). The uncanny valley (K. F. MacDorman and Norri Kageki, Trans.). *IEEE Robot. Automat.* 19, 98–100. doi: 10.1109/MRA.2012.2192811

Morris, J. D., Bradley, M., Sutherland, J., and Wei, L. (1993). Assessing cross- cultural transferability of standardized global advertising: an emotional response approach. *Paper to be Presented at National Conference of the Association in Journalism and Mass Communications,* Kansas City.

Olofsson, J. K., Nordin, S., Sequeira, H., and Polich, J. (2008). Affective picture processing: an integrative review of ERP findings. *Biol. Psychol.* 77, 247–265. doi: 10.1016/j.biopsycho.2007.11.006

Osgood, C, Suci, G., and and. Tannenbaum, P. (1957). *The Measurement of Meaning.* Urbana: University of Illinois.

Ostrom, T. M., Carpenter, S. L., Sedikides, C., and Li, F. (1993). Differential processing of in-group and out-group information. *J. Pers. Soc. Psychol.* 64, 21–34. doi: 10.1037/0022-3514.64.1.21

Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions.* Oxford: Oxford University Press.

Posner, J., Russell, J. A., and Peterson, B. S. (2005). The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* 17, 715–734. doi: 10.1017/S0954579405050340

Ramey, C. H. (2005). The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. *Paper Presented at the Proceedings of Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots,* Tsukuba.

Reis, H. T., and Gable, S. L. (2003). "Toward a positive psychology of relationships," in *Flourishing: The Positive Person and the Good Life,* eds C. L. Keyes and J. Haidt (Washington, DC: American Psychological Association), 129–159. doi: 10.1037/10594-006

Rhodes, G., Locke, V., Ewing, L., and Evangelista, E. (2009). Race coding and the other-race effect in face recognition. *Perception* 38, 232–241. doi: 10.1068/p6110

Robinson, M. D., and Compton, R. J. (2006). The automaticity of affective reactions: stimulus valence, arousal, and lateral spatial attention. *Soc. Cogn.* 24, 469–495. doi: 10.1521/soco.2006.24.4.469

Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714

Russell, J. A. (2003). Introduction: the return of pleasure. *Cogn. Emot.* 17, 161–165. doi: 10.1080/02699930302293

Schlossberg, H. (1954). Three dimensions of emotion. *Psychol. Rev.* 61, 81–88. doi: 10.1037/h0054570

Schoenherr, J. R., and Burleigh, T. J. (2015). Uncanny sociocultural categories. *Front. Psychol.* 5:1456. doi: 10.3389/fpsyg.2014.01456

Schupp, H. T., Cuthbert, B. N., Bradley, M. M., Cacioppo, J. T., Ito, T., and Lang, P. J. (2000). Affective picture processing: the late positive potential is modulated by motivational relevance. *Psychophysiology* 37, 257–261. doi: 10.1111/1469-8986.3720257

Schupp, H. T., Junghofer, M., Weike, A. I., and Hamm, A. O. (2004). The selective processing of briefly presented affective pictures: an ERP analysis. *Psychophysiology* 41, 441–449. doi: 10.1111/j.1469-8986.2004.00174.x

Schwarz, N. (2007). Attitude construction: evaluation in context. *Soc. Cogn.* 25, 638–656. doi: 10.1521/soco.2007.25.5.638

Schwarz, N., and Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: informative and directive functions of affective states. *J. Pers. Soc. Psychol.* 45, 513–523. doi: 10.1037/0022-3514.45.3.513

Schwarz, N., and Clore, G. L. (2006). "Feelings and phenomenal experiences," in *Social Psychology: Handbook of Basic Principles,* 2nd Edn, eds A. Kruglanski and E. T. Higgins (New York: Guilford), 385–407.

Schyns, P. G., and Murphy, G. L. (1994). "The ontogeny of part representation in object concepts," in *The Psychology of Learning and Motivation*, ed. D. L. Medin (New York: Academic Press), 305–349.

Seyama, J., and Nagayama, R. S. (2007).The uncanny valley: effect of realism on the impression of artificial human faces. *Pres. Teleop. Virt. Environ.* 16, 337–351. doi: 10.1162/pres.16.4.337

Seyama, J., and Nagayama, R. S. (2009). Probing the uncanny valley with the eye size aftereffect. *Pres. Teleop. Virt.* 18, 321–339. doi: 10.1162/pres.18.5.321

Smith, C. A., and Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *J. Pers. Soc. Psychol.* 48, 813–838. doi: 10.1037/0022-3514.48.4.813

Spielberger, C. D., Gorsuch, R. L., and Lushene, R. E. (1970). *Manual for the State - Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Tinwell, A. (2009). "The uncanny as usability obstacle," in *Proceedings of the "Online Communities and Social Computing" Workshop, HCI International 2009*, Vol. 12 (San Diego, CA: Springer).

Tinwell, A., and Grimshaw, M. (2009). "Bridging the uncanny: an impossible traverse?" in *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era,* eds O. Sotamaa, A. Lugmayr, H. Franssila, P. Näränen, and J. Vanhala (Tampere: ACM), 66–73. doi: 10.1145/1621841.1621855

Tinwell, A., Grimshaw, M., and Williams, A. (2010). Uncanny behaviour in survival horror games. *J. Gam. Virt. Worlds* 2, 3–25. doi: 10.1386/jgvw.2.1.3_1

Todorov, A., Pakrashi, M., and Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Soc. Cogn.* 27, 813–833. doi: 10.1521/soco.2009.27.6.813

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Q. J. Exp. Psychol. A* 43, 161–204. doi: 10.1080/14640749108400966

van Boxtel, A. (2010). "Facial EMG as a tool for inferring affective states," in *Proceedings of Measuring Behavior*, eds A. J. Spink, F. Grieco, O. Krips, L. Loijens, L. Noldus, and P. Zimmerman (Wageningen: Noldus Information technology), 104–108.

Weinberg, A., Hilgard, J., Bartholow, B. D., and Hajcak, G. (2012). Emotional targets: evaluative categorization as a function of context and content. *Int. J. Psychophysiol.* 84, 149–154. doi: 10.1016/j.ijpsycho.2012.01.023

Weyers, P., Mühlberger, A., Hefele, C., and Pauli, P. (2006). Electromyografic responses to static and dynamic avatar emotional facial expressions. *Psychophysiology* 43, 450–453. doi: 10.1111/j.1469-8986.2006.00451.x

Weyers, P., Mühlberger, A., Kund, A., Hess, U., and Pauli, P. (2009). Modulation of facial reactions to avatar emotional faces by nonconscious competition priming. *Psychophysiology* 46, 328–335. doi: 10.1111/j.1469-8986.2008.00771.x

Winkielman, P., Schwarz, N., Fazendeiro, T., and Reber, R. (2003). "The hedonic marking of processing fluency: Implications for evaluative judgment," in *The Psychology of Evaluation: Affective processes in Cognition and Emotion,* ed. J. M. K. C. Klauer (Mahwah, NJ: Lawrence Erlbaum), 189–217.

Wu, L., Winkler, M. H., Andreatta, M., Hajcak, G., and Pauli, P. (2012). Appraisal frames of pleasant and unpleasant pictures alter emotional responses as reflected in self-report and facial electromyographic activity. *Int. J. Psychophysiol.* 85, 224–229. doi: 10.1016/j.ijpsycho.2012.04.010

Yamada, Y., Kawabe, T., and Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *Japanese Psychol. Res.* 55, 20–32. doi: 10.1111/j.1468-5884.2012.00538.x.

Yen, N. S., Chen, K. H., and Liu, E. H. (2010). Emotional modulation of the late positive potential (LPP) generalizes to Chinese individuals. *Int. J. Psychophysiol.* 75, 319–325. doi: 10.1016/j.ijpsycho.2009.12.014

Yik, M. S. M., Russell, J. A., and Barrett, L. F. (1999). Structure of self-reported current affect: integration and beyond. *J. Pers. Soc. Psychol.* 77, 600–619. doi: 10.1037/0022-3514.77.3.600

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *J. Pers. Soc. Psychol. Monogr.* 9(Pt 2), 1–27. doi: 10.1037/h0025848

Zajonc, R. B. (1998). "Emotions," in *The Handbook of Social Psychology,* 4th Edn, eds D. T. Gilbert, S. T. Fiske, and G. Lindzey (Boston: McGraw-Hill), 591–632.

Zajonc, R. B. (2001). Mere exposure: a gateway to the subliminal. *Curr. Dir. Psychol. Sci.* 10, 224–228. doi: 10.1111/1467-8721.00154

Zlotowski, J., Proudfoot, D., and Bartneck, C. (2013). "More human than human: does the uncanny curve really matter?" in *Proceedings of the HRI2013 Workshop on Design of Humanlikeness in HRI from uncanny valley to minimal design.* Tokyo. 7–13.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 19 November 2014
doi: 10.3389/fpsyg.2014.01219

Perceptual discrimination difficulty and familiarity in the Uncanny Valley: more like a "Happy Valley"

Marcus Cheetham [1,2]*, ***Pascal Suter*** [1] ***and Lutz Jancke*** [1]

[1] *Department of Neuropsychology, University of Zurich, Zurich, Switzerland*
[2] *Department of Psychology, Nungin University, Seoul, South Korea*

Edited by:
Emmanuel Pothos, City University London, UK

Reviewed by:
Sabrina Golonka, Leeds Metropolitan University, UK
Rosemary A. Cowell, University of Massachusetts Amherst, USA

***Correspondence:**
Marcus Cheetham, Department of Neuropsychology, University of Zurich, Binzmühlestrasse 14/Box 25, CH-8050 Zurich, Switzerland
e-mail: m.cheetham@psychologie.uzh.ch

The *Uncanny Valley Hypothesis* (*UVH*) predicts that greater difficulty perceptually discriminating between categorically ambiguous human and humanlike characters (e.g., highly realistic robot) evokes negatively valenced (i.e., uncanny) affect. An *ABX perceptual discrimination task* and signal detection analysis was used to examine the profile of *perceptual discrimination* (*PD*) difficulty along the UVH' *dimension of human likeness* (*DHL*). This was represented using avatar-to-human morph continua. Rejecting the implicitly assumed profile of PD difficulty underlying the UVH' prediction, Experiment 1 showed that PD difficulty was reduced for categorically ambiguous faces but, notably, enhanced for human faces. Rejecting the UVH' predicted relationship between PD difficulty and negative affect (assessed in terms of the UVH' *familiarity* dimension), Experiment 2 demonstrated that greater PD difficulty correlates with more positively valenced affect. Critically, this effect was strongest for the ambiguous faces, suggesting a correlative relationship between PD difficulty and feelings of familiarity more consistent with the metaphor *happy valley*. This relationship is also consistent with a *fluency amplification* instead of the hitherto proposed *hedonic fluency* account of affect along the DHL. Experiment 3 found no evidence that the asymmetry in the profile of PD along the DHL is attributable to a *differential processing bias* (cf. *other-race effect*), i.e., processing avatars at a category level but human faces at an individual level. In conclusion, the present data for static faces show clear effects that, however, strongly challenge the UVH' implicitly assumed profile of PD difficulty along the DHL and the predicted relationship between this and feelings of familiarity.

Keywords: perceptual discrimination, categorical perception, categorization, uncanny valley, human likeness, other-race effect, processing fluency, mere exposure

INTRODUCTION

Progress in robotics and computer graphics in simulating human appearance and behavior to high degrees of realism has fuelled research interest in the *Uncanny Valley Hypothesis* (*UVH*) (Mori, 1970). The UVH predicts that perceptual difficulty discriminating between highly realistic humanlike objects and characters (e.g., robot, prosthetic hand) and their human equivalent will evoke an unpleasant affective state. This state is described as one of feelings of personal disquiet, strangeness and the uncanny. These feelings are conjectured to occur at the point of realism along the UVH' *dimension of human likeness* (*DHL*) at which the attribution of objects and characters to the human or nonhuman category is subject to greatest ambiguity (i.e., the "valley" in **Figure 1**). Studies to date have not provided a consistent picture in favor of this uncanny effect, but this field of research is still in its infancy (e.g., Hanson, 2006; MacDorman, 2006). Possibly for this reason, almost no attention has been given to determining where along the DHL there is greater difficulty in *perceptual discrimination* (*PD*) (Looser and Wheatley, 2010; Cheetham et al., 2011) and to whether greater PD difficulty does relate to an increase in negative affective experience.

The UVH' prediction is based on the implicit assumption that PD difficulty is greatest at or near the point along the DHL at which there is greatest categorization ambiguity (i.e., the category boundary). There are two potential problems with this assumption. The first is that it conflicts with the general consensus that there is normally less PD difficulty at or near the category boundary compared with other regions of a perceptual dimension like the DHL (e.g., Harnad, 1987). The second is that PD difficulty might actually be most pronounced for categorically unambiguous human stimuli compared with other stimuli along the DHL. The potential impact of these two problems on the UVH is apparent in the following thought experiment. If we assume that PD difficulty is in fact attenuated at the category boundary and enhanced for human category exemplars (as tested in Experiment 1 of the present study) but that the UVH's prediction is otherwise correct (i.e., a positive relationship between PD difficulty and negative affect), it follows that greater negative affect should be experienced for objects and characters at the human category end of the DHL. This conclusion is, however, difficult to reconcile with the very idea that Mori sought to express in the UVH and more generally with reports of increased positive

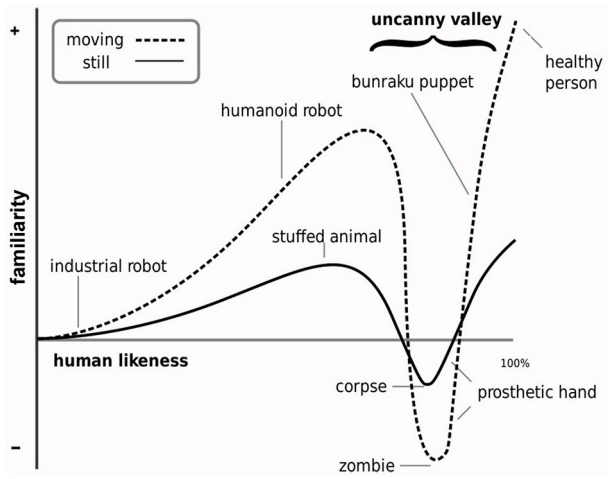


FIGURE 1 | Illustration of the Uncanny Valley Hypothesis. The uncanny valley hypothesis proposes a non-linear relationship between affective experience and physical humanlike realism in appearance and motion (the non-linearity is more pronounced for motion). The key prediction of the hypothesis is that a high degree of human likeness will evoke a sharp negative peak (valley) in affective experience (i.e., along the *familiarity dimension*). This valley is characterized by feelings of strangeness (and the uncanny). The valley occurs at the point along the *dimension of human likeness* at which objects are categorically most ambiguous (illustration adapted from MacDorman, 2005).

affect for human compared with nonhuman faces (e.g., Looser and Wheatley, 2010), unless it is assumed that the direction of the UVH' conjectured relationship between PD difficulty and affect is also incorrect (as tested in Experiment 2 of the present study).

These potential problems with the implicitly assumed distribution of PD difficulty along the DHL can be considered in terms of the literature on *categorical perception* (CP, for CP see e.g., Goldstone and Hendrickson, 2010; for the similar *perceptual magnet effect*, see Kuhl, 1991). CP refers to the phenomenon that the cognitive representation of *psychological similarity space* (such as along a perceptual dimension like the DHL) can be selectively deformed (Livingston et al., 1998). Deformation (or warping) of psychological similarity space is evident when, relative to a baseline of comparison, physical differences between stimuli within a category are subjectively perceived to be more similar (i.e., less discriminable) than equally spaced physical differences between stimuli from two different categories that are subjectively perceived to be less similar (i.e., more discriminable).

Many studies of CP have focused on facial processing. Studies of CP such as for famous faces (Beale and Keil, 1995), unfamiliar faces (Levin and Beale, 2000), facial expressions (Ectoff and Magee, 1992) and faces of different gender (Bülthoff and Newell, 2000) reveal a relatively symmetrical pattern of warping and PD performance along facial continua. This pattern is characterized by enhanced discriminability of stimuli (i.e., less PD difficulty) at the category boundary (rather than greater PD difficulty as assumed in the UVH) and similarly attenuated discriminability of stimuli (i.e., greater PD difficulty) within both categories. The similarly attenuated PD within the categories likely relates to the assumption in studies of CP (such as those in the preceding) that there is comparable (or symmetrical) category knowledge, categorization experience, and processing of continua endpoints from which morph continua are generated.

In contrast to this symmetry, the UVH was originally formulated on the basis of (informal) observation of individuals with extensive everyday experience processing human others but comparably little perceptual and categorization experience processing humanlike robotic characters. This implicit assumption in the UVH that categorization experience is asymmetrical for human compared with nonhuman others is also implicit in most uncanny-related studies (e.g., Yamada et al., 2013). These studies have typically examined participants who have everyday expertise in facial processing of human category exemplars (see Diamond and Carey, 1977; Tanaka, 2001) but, by virtue of the innovative nature of avatar and robot research and design and of the methods used to generate experimental stimuli, comparatively little if any such experience processing the subtle perceptual manipulations of and differences in human likeness between the nonhuman stimuli under investigation.

It is conceivable that differential experience in perceptual and category information processing will influence perceptual sensitivities and PD difficulty along the DHL. (Gibson, 1991; Hall, 1991; Goldstone, 1994; Harnad, 1987; Sigala et al., 2011). For example, compared with PD performance before training (using continua for which symmetrical knowledge of continua endpoints can be assumed), categorization experience with novel continua based on line drawings of fictitious animals (Livingston et al., 1998), with natural unfamiliar faces (Kikutani et al., 2008, 2010) and with faces of identical twins (Stevenage, 1998) is reflected in greater PD difficulty for within-category stimuli and lesser PD difficulty for the between category stimuli that straddle the category boundary. It would be consistent with such findings that asymmetry in categorization experience with human faces (for which there is everyday expertise due to a history of normal social interaction) compared with novel nonhuman faces (for which there is comparatively little or no such expertise) is reflected in a corresponding asymmetry in PD performance along the DHL. This would mean greater PD difficulty for within-category human stimuli compared with lesser PD difficulty for within-category nonhuman stimuli.

In the first of three experiments, we tested whether the distribution of PD difficulty along the DHL implicitly assumed in the UVH is correct. Considering the influence of categorization experience on CP and psychological similarity space reported in the preceding studies, we anticipated, firstly, that faces within the human category would generally be more difficult to discriminate compared with those closest to or at the category boundary. Second, we anticipated that faces within the human category would generally be more difficult to discriminate compared with those within the nonhuman category. To examine this, we delineated the profile of PD performance for morphed faces drawn from morph continua representing the DHL. The continua were generated from avatar (i.e., computer-generated characters) and human parent faces. The morphed faces were presented in an *ABX PD task* (Liberman et al., 1957; this task is described in detail

in Section Design and procedure). Campbell et al. (1997) used the ABX task to investigate CP along other dimensions of human likeness and showed that this task is sensitive to differences in perceptual processing between human and nonhuman faces. Signal detection analysis was used to assess discrimination sensitivity. A *two-alternative forced choice categorization task* (described in detail in Section Design and procedure) was conducted after the ABX task in order to define the profile of categorisation ambiguity and the location of the category boundary along the continua. The second experiment replicated the findings of the first experiment. In the second experiment, we tested the UVH' predicted relationship between increased PD difficulty and negative affective experience. In the third experiment, we explored the possibility that the asymmetry in PD difficulty along the DHL reported in Experiments 1 and 2 might be attributable to a *differential processing bias*. This bias means that avatars might be processed at a category level and human faces at an exemplar level, this resulting in differences in PD performance between avatar and human faces of the DHL.

STUDY 1: ABX PERCEPTUAL DISCRIMINATION AND FORCED CHOICE CATEGORIZATION TASKS

MATERIALS AND METHODS

Participants

Healthy adult volunteers ($N = 49$, 29 female, mean age 21.8 years; range 19–25 years) with no record of neurological or psychiatric illness and no current medication use were recruited for the study. All study participants were students of the University of Zurich, native or fluent speakers of Swiss or Standard German, and consistently right-handed, as assessed with self-rating scales (Annett, 1970). Each participant confirmed after completion of the experiment having had no previous experience designing or modifying computer-generated characters as for example in *virtual reality* (*VR*) role-playing games, second life, or VR environments or using such environments (e.g., for psychotherapy, rehabilitation, training, e-commerce or virtual reality-based research) and explicitly no previous experience (e.g., in video games) with the kind of highly humanlike characters and manipulations of human likeness presented in the current study. At debriefing, one participant reported uncertainty about the correct use of the response buttons and 3 others about the meaning of the label "avatar" in the forced choice categorization task. Analyses with and without the data of these four participants had no impact on the pattern of findings. The findings are reported on the basis of the complete data set. Written informed consent was obtained before participation according to the guidelines of the Declaration of Helsinki. Each volunteer received 20 Swiss Francs for participation. The study and all procedures and consent forms were approved by the Ethics Committee of the University of Zurich.

Stimuli

Morph continua were generated to represent the DHL, using the software Fantamorph® (Abrosoft http://www.abrosoft.com). These were produced in the same way as in previous studies (for an example continuum, see e.g., Cheetham et al., 2011). Eight color photographic images of natural faces and 8 color images of avatar faces were used as parent faces to produce 8 morph continua. The selection of continua was based on previous pilot testing to ensure like performance across continua (i.e., same morph position of the category boundary and shape of the response function). Each continuum comprised 11 different morphed images from the avatar endpoint (number 1) to the human endpoint (number 11), each morph being separated by an increment of 10% in physical difference (see **Figure 2B**). All parent faces were male, indistinctive, presented with full face, frontal view, direct gaze, neutral expression and no other salient features such as facial hair and jewelry. Avatars were generated with the modeling suite Poser 7® (Smith Micro Software, http://www.smithmicro.com) for detailed adjustment of facial geometry and texture (e.g., age and configural cues) to closely match the corresponding human face. Matching aimed to minimize perception of biological motion due to quick successive presentation of morphs (Schultz and Pilz, 2009) and to ensure perception of faces in the two-step procedure of the ABX as having the same identity. Adobe Photoshop 7.0® (http://www.adobe.com) was used for image editing. Before morphing, the external features of each parent face were masked with an elliptic form and black background (96 dpi and 560 × 650 pixels), and contrast levels, overall brightness and skin tone of the parent faces of each continuum were adjusted to match.

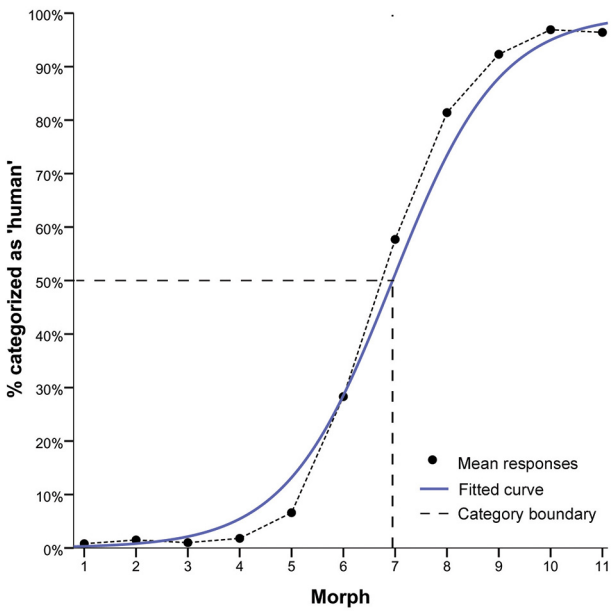


FIGURE 2 | Results of the forced choice classification task. Mean responses are depicted in terms of % of "human" responses. The mean grand average across all continua (continuous blue line), fitted logistic curve based on the grand mean (black line), and the category boundary (dashed gray line) are shown. The category boundary indicates the point of maximum uncertainty of 50% in categorisation judgements along the continua. The logistic-shaped curve shows a lower and upper asymptote of avatar and human categorisation responses and a step-like response function consistent with the presence of a category boundary. Morph M7 shows the greatest categorisation ambiguity.

Design and procedure

All participants were tested individually by a research assistant blind to the purpose and hypotheses of the study. Following established procedure (Newell and Bulthoff, 2002), the PD ABX task was conducted first followed by the two-alternative forced-choice categorization task. The experiment lasted approximately 40 min, with a short break between the discrimination and categorization tasks.

Perceptual discrimination ABX task. The UVH does not suggest how DP difficulty should be operationalised and tested. For PD, the ABX discrimination task was used (Liberman et al., 1957; Harnad, 1987). This entails presentation of trials in which pairs of different face stimuli (A and B) are followed by a second presentation of either A or B as the target stimulus X. Participants are required to view all three images and respond by button press to indicate whether A or B is identical to (i.e., the same as) X. A 2-step discrimination procedure was applied so that stimulus B differed in physical distance along the continuum from stimulus A by two steps (i.e., 1–3, 2–4, 3–5, etc.). To counterbalance the sequence of face pairs, each pair was presented four times, once in each of the possible combinations (i.e., AB-A, BA-B, AB-B, BA-A). Both faces of each presented pair were always drawn from the same continuum in which they were originally morphed. The presentation of face pairs was pseudo-randomized so that no trails using face pairs from corresponding morph positions of other continua were presented in sequence.

Written instructions were presented on the screen before commencement of the experiment. Participants performed a pre-test of 5 trials (using stimuli drawn at random from continua that were not included in the main test) to ensure comprehension of the instructions and correct use of the response buttons. The background on the monitor was always black. Stimuli A and B were presented for 750 ms immediately followed by stimulus X, which remained on screen until the response was made or till time-out at 4 s. The inter-trail interval was 1500 ms. Response accuracy and *response time* (*RT*) were measured for each trial, including the practice trials.

The ABX task (and forced-choice classification task described in the next section) was conducted in a sound attenuated and light-dimmed room, and morph stimuli were presented on a LCD monitor (1280 × 1024 resolution, 60 Hz refresh rate), using Presentation® software (Version 14.1, www.neurobs.com). The stimuli (400 × 500 pixels) were presented at a viewing distance of 62 cm.

Two-alternative forced-choice categorization task. The same stimuli presented in the ABX task were presented in a two-alternative forced-choice categorization task. This task commenced with the presentation of written instructions. Subsequently, participants performed a practice pre-test of 5 trials, using the same stimuli used in the pre-trials of the ABX task. Having ensured task comprehension and correct use of the response buttons, the participant initiated testing by pressing a button. The forced-choice categorization task normally follows the PD task in order to minimize the potential influence of labeling on discrimination performance (Newell and Bulthoff, 2002). To minimize this further, the labels "avatar" and "human" were first used during task instruction for the forced choice task. The background on the monitor was always black. All morph stimuli were presented twice, individually, centrally, and in random order with the constraint that stimuli from corresponding morph positions of other continua were not presented in sequence. Each trial began with the presentation of a fixation point for 500 ms (participants were required to maintain fixation), followed by a morph image for 750 ms. The participant was asked to identify the stimulus quickly and accurately as either an avatar or human by pressing one of two response keys. A black screen with fixation point remained after presentation of the morph image until the participant pressed the response key, after which a blank black screen without fixation cross remained for 1500 ms until the next trial began.

All data analyses were performed using SPSS version 21.0 (http://www.ibm.com). MATLAB 2006b (http://www.mathworks.ch) was used to implement the Palamedes routines (Prins and Kingdom, 2009) for signal detection analysis of data from the ABX task.

RESULTS

The response data for avatar vs. human category judgments in the forced choice categorization task were analyzed (Section Forced choice categorization task: Responses, logistic function, and category boundary) to determine the choice of categorically ambiguous and unambiguous morphs for use in the analyses of PD performance (Section Forced choice categorization task: Response times).

Forced choice categorization task: responses, logistic function, and category boundary

The slope of the categorization response function was used to summarize the category judgments by fitting logistic function models to the data of each participant across continua. The parameter estimates derived from each model were entered in analyses of logistic function of categorization responses and of the category boundary.

For the logistic function of categorization responses, the parameter estimates were tested against zero in a one-sample *t*-test. The result showed a highly significant logistic component [$t_{(48)} = 44.31$, $p > 0.001$] consistent with the presence of a category boundary (Harnad, 1987) (see **Figure 2**).

To compute the value of the category boundary (i.e., $y = 0.5$: -ln[β0]/ln[β1]), the parameter estimates β0 and β1 of each participant's logistic function model were used. The mean category boundary value was $M = 6.95$. This value indicates the actual morph position along the continua that corresponds with the ordinate midpoint between the lower and upper asymptotes, that is, the point of maximum uncertainty of 50% in categorization judgments. Across continua, morph M7 is closest to this boundary (**Figure 2**; see also Supplemental Figures 1, 3, 5 for the results of the forced choice categorization task of each experiment with error bars).

To show this profile of high and low ambiguity in categorization judgments more clearly, we tested for differences in category decisions between the unambiguous avatar (i.e., M3,

M4, M5) and human faces (i.e., M9, M10, M11) and the most ambiguous faces (i.e., M7). This choice of morphs permitted control for physical morph distance along continua between the ambiguous M7 and the unambiguous avatar and human faces. A one-way *repeated measures of analysis of variance* (*RM-ANOVA*) was performed on the dependent variable mean "categorization" response of each participant across continua, using the factor "morph" position (3 levels: "M3, M4, M5," "M7," "M9, M10, M11"). Greenhouse-Geisser adjustment was applied to correct the degrees of freedom for violation of the sphericity assumption (and applied as appropriate in all subsequent analyses). This analysis showed a highly significant effect for morph position, $F_{(1.27,\ 58.93)} = 455.26$, $p < 0.001$. Mean categorization difficulty for M7 was $M = 0.58$ ($SE = 0.04$), while that for the human faces was $M = 94.52$ ($SE = 0.01$) and for avatar faces $M = 4.02$ ($SE = 0.01$) (see **Figure 3**).

Forced choice categorization task: response times

Differences in category ambiguity, as indicated by the logistic-shaped response function, are likely to be reflected in different *RT* for category judgments. Before data analysis, short RT latencies of less than 100 ms were excluded. RT data for long latency outliers were screened by z-standardizing and filtering out data points using $z = 3$ as a cut-off score (Van Selst and Jolicoeur, 1994). Analyses were conducted with and without outliers. These analyses produced the same pattern of results. The findings are therefore reported for the complete data set. Confirming RT differences in category decision difficulty, a one-way RM-ANOVA with morph position (11 levels: M1-M11) and RT as the dependent variable showed a main effect for morph position, $F_{(4.58,\ 215.09)} = 41.23$, $p < 0.001$.

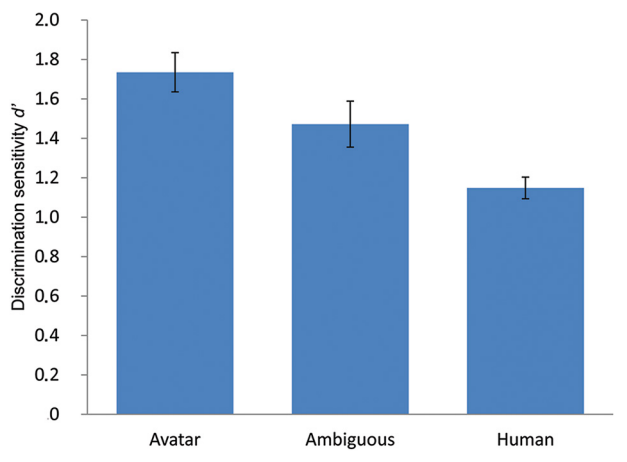


FIGURE 3 | Results of the ABX perceptual discrimination task. The figure depicts mean discrimination sensitivity d' in the ABX perceptual discrimination task for unambiguous avatar and human and highly ambiguous faces. The profile of d' shows a marked asymmetry along the continua, meaning that unambiguous avatar (and the ambiguous faces) are perceived as more dissimilar than equally spaced human faces. Contrary to the implicit assumption in the UVH, perceptual discrimination difficulty is greatest for human faces. The error bars indicate 1 *SE* ($N = 49$).

The longest response latencies would be expected to correspond with the morph position closest to the category boundary, that is, at M7 (see Supplemental Figure 2). But inspection of Supplemental Figure 2 indicates that RT for M6 and M7 are similarly long. The tests of planned within-subject contrasts in the preceding analysis showed no significant difference between M6 and M7 in RT. Given that M7 and the category boundary are so closely aligned, the following analysis compared ambiguity at M7 with the unambiguous avatar (i.e., M3, M4, M5) and human faces (i.e., M9, M10, M11), but a re-run of the same analysis using the aggregate mean of M6 and M7 instead of just M7 produced the same pattern of results. A one-way RM-ANOVA analysis with "morph" positions (3 levels: "M3, M4, M5," "M7," "M9, M10, M11") and RT in ms as dependent variable was conducted. The analysis showed a highly significant effect for morph position [$F_{(2.96,\ 58.93)} = 45.22$, $p < 0.001$]. Pre-planned contrasts showed that RT was longer significantly longer for human ($M = 1073$, $SE = 44$) than for avatar faces ($M = 898$, $SE = 33$), $F_{(1,\ 48)} = 15.72$, $p > 0.001$, and that RT for M7 ($M = 1348$, $SD = 60$) differed highly significantly from RT for the other avatar and human morph positions ($M = 928$, $SD = 0.19$), $F_{(1,\ 48)} = 67.49$, $p < 0.001$.

ABX Perceptual discrimination task

Differences in the ability to perceptually discriminate between pairs of morphs (M6-M8) straddling the ambiguous M7 and between pairs of unambiguous morphs within the avatar (M3-M5, M4-M6) and human (M8-M10, M9-M11) face categories were tested. This choice of avatar and human morph pairs ensured control for the physical morph distance along the continua between the ambiguous and unambiguous faces. The mean value of PD was compared in a one-way RM-ANOVA with factor morph position (3 levels: "M3-M5, M4-M6," "M6-M8," "M8-M10, M9-M11") using d' as dependent variable (Best et al., 1981). d' is used a measure of discrimination performance derived from Signal Detection Theory (e.g., Macmillan and Creelman, 2005) that takes effects of response bias (c) into account. This measure is used instead of the percentage of correct different responses to different pairs (Francis and Ciocca, 2003). A differencing model was applied to compute d' because this is considered to best reflect the decision strategy used in the ABX task (Pierce and Gilbert, 1958; Hautus and Meng, 2001; Macmillan and Creelman, 2005).

This analysis showed a significant effect for morph pair position, $F_{(2,\ 96)} = 14.68$, $p < 0.001$. Tests of planned within-subject contrasts showed that PD of faces within the avatar category ($M = 1.74$, $SE = 0.1$) was significantly greater than that of ambiguous faces at the category boundary ($M = 1.47$, $SE = 0.12$) [$F_{(1,48)} = 5.59$, $p = 0.022$] and of faces within the human category ($M = 1.15$, $SE = 0.05$), $F_{(1,\ 48)} = 38.54$, $p < 0.001$. PD of ambiguous faces was significantly greater than that of faces within the human category, $F_{(1,\ 48)} = 7.5$, $p = 0.009$.

A one-way RM-ANOVA with "morph position" (11 levels) and c as the dependent variable for response bias showed no significant differences for c.

DISCUSSION

The data confirm that there are differences in PD difficulty as a function of human likeness along the DHL. But the pattern of PD is entirely different than that implicitly assumed in the UVH. Firstly, and as expected on the basis of previous studies of CP, PD of faces at the category boundary is enhanced compared with PD of within-category human faces. Second, PD of within-category avatars is also enhanced compared with PD of within-category human faces, thus supporting the suggestion that PD performance along the DHL might be asymmetrical.

Given that the UVH predicts enhanced negative affective experience as a function of enhanced PD difficulty, these findings would mean—assuming that the UVH is otherwise correct—that human faces should evoke more negative affect compared with ambiguous faces and unambiguous avatar faces. This is clearly inconsistent with the idea that Mori sought to convey in his graphical representation of his hypothesis, and the available evidence from uncanny-related research suggests that enhanced feelings of strangeness for human category exemplars is highly unlikely. Self-ratings of comparably well-controlled morph continua show that positive ratings (e.g., pleasantness) increase with greater human likeness (e.g., Looser and Wheatley, 2010).

In a second experiment, we tested whether there is nevertheless evidence in favor of the UVH' prediction that enhanced PD difficulty is associated with greater negative affective experience. The UVH conceptualizes affective experience as *shinwakan*, an ambiguous Japanese neologism that Mori used to describe the positive and negative character of affective experience of human-like objects. There have been various renderings of shinwakan's meaning in uncanny-related research, including comfort level, familiarity, eeriness, pleasantness, likability, empathy and affinity (e.g., MacDorman and Ishiguro, 2006; Bartneck et al., 2007; Seyama and Nagayama, 2007; Green et al., 2008; Tinwell et al., 2011; Dill et al., 2012; Mori, 2012; MacDorman et al., 2013; Burleigh et al., 2013; see also Ho and MacDorman, 2010). To examine affective experience, we used an *ad hoc* self-rating scale based on the UVH' bi-polar dimension of *familiarity* (i.e., feelings of familiarity vs. strangeness). Familiarity was selected because this rendering of shinwakan has been used frequently in research, it is most often used to denote the affective dimension of the UVH in its illustration (see **Figure 1**), and because it arguably best captures the apparent meaning of shinwakan that Mori sought to convey in the UVH's description. Clearly, there are alternative approaches to examining affective experience of human like objects and characters based on well-validated dimensions of affective experience and measures of these. The aim of this experiment was to test affective experience as conceptualized in the UVH in relation to PD difficulty.

EXPERIMENT 2

The materials, methods and analyses in Experiment 2 were identical to those in Experiment 1, with two exceptions. Firstly, the presented morphs were drawn from continua that were generated anew. This was done by switching the source image (i.e., avatar) and destination image (i.e., human) for morphing in Experiment 1 so that the human was now the source and the avatar the destination image. The continua were then re-morphed, and the morphs were labeled M1 (avatar) to M11 (human) as in Experiment 1. The reason for switching the source and destination images and of re-morphing the stimuli was to exclude the possibility that the strong asymmetry in PD performance in Experiment 1 was simply a systematic artifact of any nonlinearity in the morphing algorithm used to generate the continua. If it was a systematic artifact, the PD data in Experiment 2 would show a similarly skewed pattern of PD along the DHL, with however enhanced PD for the human instead of the avatar faces. Second, participants performed the ABX task followed this time by the self-rating task, in which to report feelings of familiarity, and only then by the two-alternative forced choice categorization task. The latter task was performed last to ensure that any effects in ratings were not biased by explicit processing of faces for forced categorization.

The UVH does not suggest how DP difficulty and feelings of familiarity should be operationalised and tested. We used our measure of discrimination sensitivity d' to indicate DP performance, as applied in Experiment 1, and, in keeping with the favored approach to date in uncanny research, we used subjective ratings to indicate feelings of familiarity in the self-rating task. The task requirements, instructions and stimulus presentation conditions of the self-rating task were identical to those described for the two-alternative forced choice categorization task in Experiment 1, with the exception that participants viewed and rated the subjective feeling of familiarity evoked by each morphed stimulus on a 5-point Likert scale. The scale ranged from very strange (1) to very familiar (5). To test the relationship between DP difficulty and feelings of familiarity we took an inter-individual differences approach. We tested whether individual variability in the ability to discriminate between a pair of morphed faces (e.g., M2-M4) predicts individual variability in self-rated feelings of familiarity for the face (e.g., M3) that the given face pair straddles. This approach assumes that there are stable individual differences in the relationship between familiarity ratings and discrimination performance. If Mori's prediction is correct, greater PD difficulty should be associated with increased feelings of strangeness (i.e., with less familiarity). This was tested.

PARTICIPANTS

A new sample of $N = 49$ volunteers (34 female, mean age 21.9 years; range 19–31 years) not involved in Experiment 1 participated in Experiment 2.

RESULTS

Forced choice categorization task: responses, logistic function, and category boundary

The parameter estimates derived from each logistic function model of each participant across continua were tested against zero in a one-sample *t*-test and showed, as in Experiment 1, a highly significant logistic component [$t_{(48)} = 27.83$, $p > 0.001$] (see **Figure 4**). Based on the parameter estimates β0 and β1, the mean category boundary value was $M = 6.6$. Across continua, the most ambiguous face morph M6 is closest to this boundary.

To show the effects of this profile of high and low ambiguity in categorization judgments more clearly, we tested for differences in category decisions between the unambiguous avatar (i.e.,

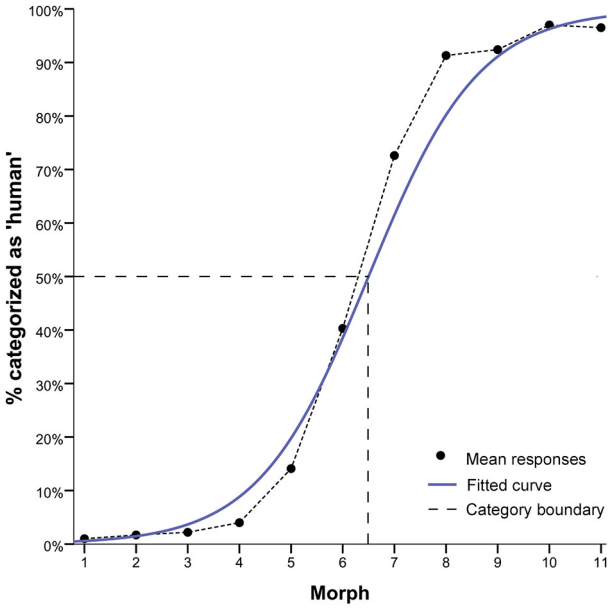


FIGURE 4 | Results of the forced choice classification task. Mean responses are depicted in terms of % of "human" responses, with the grand average across all continua (continuous blue line), the fitted logistic curve based on the grand mean (black line), and the category boundary (dashed gray line) to indicate the point of maximum uncertainty of 50% in categorisation judgements along the continua. Results indicate a step-like response function consistent with the presence of a category boundary. Morph M6 shows the greatest categorisation ambiguity.

M2, M3, M4) and human faces (i.e., M8, M9, M10) and the most ambiguous faces (i.e., M6). Consistent with the approach in Experiment 1, the choice of morphs permitted control for physical morph distance along continua between M6 at the category boundary and the avatar and human faces. A one-way *RM-ANOVA* was performed on the dependent variable mean "categorization" response of each participant across continua, using the factor "morph" position (3 levels: "M2, M3, M4," "M6," "M8, M9, M10"). This analysis showed a highly significant effect for morph position [$F_{(1.22,\ 58.52)} = 483.72$, $p < 0.001$]. Categorization difficulty for M6 was closest to chance level of 50% ($M = 40.31$; $SE = 3.73$), while that for the human faces was $M = 93.58$ ($SE = 1.13$) and for avatar faces $M = 2.63$ ($SE = 0.44$) (see **Figure 4**).

Forced choice categorization task: response times

We verified whether differences in category ambiguity are reflected in the RT for category judgments. Data were screened for outliers as in Experiment 1 and analyses conducted with and without these. These analyses produced the same pattern of results for which reason the findings for the complete data set are reported. Confirming RT differences in category decision difficulty, a one-way RM-ANOVA with morph position (11 levels: M1-M11) and RT as the dependent variable showed a main effect for morph position, $F_{(4.58,\ 220.22)} = 39.03$, $p < 0.001$.

Inspection of the RT data (see Supplemental Figure 4) indicates that the longest response latencies correspond with the most ambiguous morph M6. A one-way RM-ANOVA analysis with "morph" positions (3 levels: "M2, M3, M4," "M6," "M8, M9, M10") and RT in ms as dependent variable was conducted. The analysis showed a highly significant effect for morph position, $F_{(2,\ 96)} = 54.99$, $p < 0.001$. Pre-planned contrasts showed that RT was longer significantly longer for human ($M = 957$, $SE = 34$) than for avatar faces ($M = 751$, $SE = 23$), $F_{(1,\ 48)} = 15.72$, $p > 0.001$, and that RT for M6 ($M = 1191$, $SD = 53$) differed highly significantly from RT for the other morph positions ($M = 851$, $SD = 0.37$), $F_{(1,\ 48)} = 65.91$, $p < 0.001$.

ABX perceptual discrimination task

An independent samples *t*-test (Experiment 1 vs. Experiment 2) using d' for each morph pair position in the ABX task (i.e., pairs M1-M3 through to M9-M11) of each participant across continua as dependent variable showed that discrimination performance for each morph pair was not significantly different between Experiments 1 and 2. The following results indicate also that the PD effects in Experiment 1 are comparable to those in Experiment 2.

Given that face morph position M6 was closest to the category boundary in Experiment 2, differences in the ability to perceptually discriminate between pairs of morphs (M5-M7) straddling the *ambiguous* M6 was compared with the ability to perceptually discriminate between unambiguous morphs within the *avatar* (M1-M3, M2-M4, M3-M5) and *human* (M7-M9, M8-M10, M9-M11) face categories. These morph pairs were selected because they straddle the morph positions M2, M3, M4, M6, M8, M9, M10 that were analyzed in the forced choice task of Experiment 2 and because this choice of pairs ensures control for physical morph distance between the ambiguous and the unambiguous human and avatar faces. The mean value of discrimination sensitivity was compared in a one-way RM-ANOVA with factor morph position (3 levels: "M1-M3, M2-M4, M3-M5," "M5-M7," "M7-M9, M8-M10, M9-M11") using d' as dependent variable.

This analysis showed a significant effect for morph pair position, $F_{(2,\ 96)} = 16.52$, $p < 0.001$ (see, **Figure 5**). Tests of planned within-subject contrasts showed that discrimination of avatar faces ($M = 1.41$, $SE = 0.08$) was significantly greater than that of faces within the human category ($M = 0.99$, $SE = 0.07$), $F_{(1,\ 48)} = 27.59$, $p > 0.001$. Discrimination of ambiguous faces at the category boundary ($M = 1.53$, $SE = 0.11$) was not significantly greater than that of faces within the avatar category ($M = 1.41$, $SE = 0.08$) [$F_{(1,\ 48)} = 1.11$, $p = 0.299$], but it was significantly greater than that of faces within the human category ($M = 0.99$, $SE = 0.07$), $F_{(1,\ 48)} = 25.76$, $p < 0.001$.

It should be noted that the most ambiguous morph was M7 in Experiment 1 and M6 in Experiment 2. This means that the choice of morph pairs for inclusion in the analyses of d' in Experiment 1 is partially different than the choice in Experiment 2. To compare Experiments 1 and 2, the one-way RM-ANOVA in Experiment 2 was re-run, using this time the same morph positions selected in Experiment 1, that is, M3-M5 and M4-M6 for avatar faces, M6-M8 for the ambiguous M7, and M8-M10 and M9-M11 for human faces. This analysis showed the same pattern of significant effects for morph pair position [$F_{(2,\ 96)} = 21.42$, $p < 0.001$] and for the tests of planned within-subject contrasts (see Supplemental Figure 6). The contrasts showed that PD of

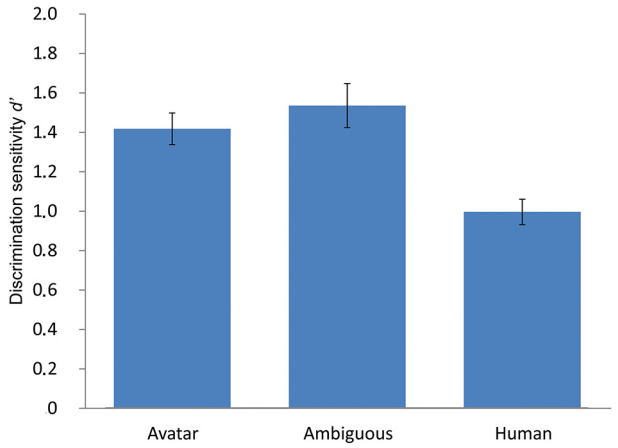


FIGURE 5 | Results of the ABX perceptual discrimination task. The figure depicts mean discrimination sensitivity *d'* in the ABX perceptual discrimination task for unambiguous avatar and human and highly ambiguously faces. The profile of discrimination sensitivity replicates that found in the first experiment. Perceptual discrimination difficulty is greatest for human faces. The error bars indicate 1 SE ($N = 49$).

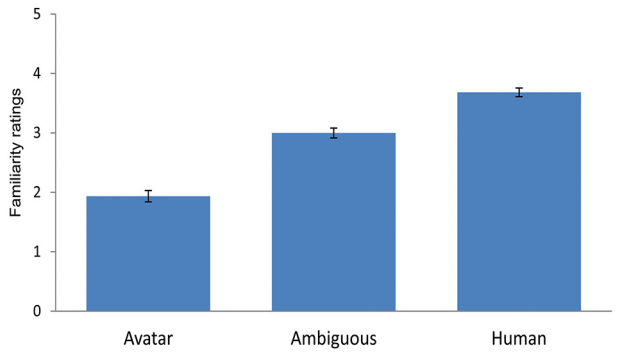


FIGURE 6 | Results of self-rating task for familiarity. Overall, the figure illustrates a general increase in self-rated feelings of strangeness with decreasing human likeness from the human end of the continua. There is no indication of the uncanny effect predicted in the UVH. The error bars indicate 1 *SE* ($N = 49$).

faces within the avatar category ($M = 1.67$, $SE = 0.11$) was significantly greater than that of ambiguous faces at the category boundary ($M = 1.34$, $SE = 0.11$) [$F_{(1,48)} = 12.87$, $p = 0.001$] and of faces within the human category ($M = 0.98$, $SE = 0.07$), $F_{(1, 48)} = 35.48$, $p < 0.001$. PD of ambiguous faces was significantly greater than for faces within the human category, $F_{(1, 48)} = 11.26$, $p = 0.002$. Taken together, these analyses are consistent in indicating asymmetry in discrimination performance along the continua.

A one-way RM ANOVA with "morph position" (11 levels) and *c* as the dependent variable for response bias showed no significant differences for *c*.

FAMILIARITY RATINGS

Differences in mean familiarity ratings between the *unambiguous* avatar (i.e., M2, M3, M4) and *human* faces (i.e., M8, M9, M10) and the most *ambiguous* faces (i.e., M6) were tested using the same morph positions as in the analysis of the forced choice categorization task in Experiment 2 (Section Forced choice categorization task: Responses, logistic function, and category boundary). A one-way RM-ANOVA with the factor *morph* position (3 levels: "M2, M3, M4," "M6," "M8, M9, M10") and the dependent variable *familiarity* rating of each participant across continua revealed a highly significant effect of morph position, $F_{(1.48,\ 70.93)} = 180.61$, $p \leq 0.001$ (see **Figure 6**). Pre-planned contrasts showed a significant difference between the avatar morphs ($M = 1.93$; $SE = 0.1$) and M6 ($M = 3$; $SE = 0.08$) $F_{(1,\ 48)} = 278.67$, $p \leq 0.001$ and between M6 and the human morphs ($M = 3.68$; $SE = 0.07$), $F_{(1,\ 48)} = 53.02$, $p \leq 0.001$, and between the avatar and human morphs, $F_{(1,\ 48)} = 27.59$, $p \leq 0.001$. Taken together, the data indicate that familiarity ratings increase negatively (i.e., greater strangeness) across the three stimulus conditions with increasing distance from the human end of the continua. This lends no support to the UVH' predicted increase in negative evaluations for the most ambiguous faces.

RELATIONSHIP BETWEEN PERCEPTUAL DISCRIMINATION AND FAMILIARITY RATINGS

The UVH predicts a positive relationship between greater PD difficulty and greater subjective experience of strangeness. To test this we examined whether individual variability in PD performance for face pairs predicts individual variability in ratings of subjective experience for the faces that the face pairs straddle. Pearson product-moment correlations were conducted using the mean data of each participant across continua of each *morph* in the familiarity rating task (i.e., "M2, M3, M4" for avatar, M6 for ambiguous, and "M8, M9, M10" for human faces) and the morph pairs that straddled these faces in the ABX task (i.e., "M1-M3, M2-M4, M3-M5" for *avatar*, M5-M7 for ambiguous, M7-M9, M8-M10, M9-M11 for human faces). Outlier detection was performed before analysis by means of boxplots. This indicated 1 outlier. After removal of this outlier, the analyses showed a highly significant (two-sided) negative correlation between PD performance and familiarity ratings for avatar faces [$r_{(48)} = -0.314$, $p = 0.03$] and for ambiguous faces [$r_{(48)} = -0.494$, $p > 0.001$]. There was no significant relationship between PD performance and familiarity ratings for human faces [$r_{(49)} = 0.088$, $p = 0.533$].

DISCUSSION

The data of Experiment 2 replicated those of Experiment 1 by showing the same pattern of PD asymmetry, that is, enhanced PD for highly ambiguous faces and highly unambiguous nonhuman faces but attenuated discrimination for highly unambiguous human faces. Based on a new sample of participants and re-morphed continua, this pattern re-affirms that the implicit assumption in the UVH, that is, greater PD difficulty in the categorically most ambiguous region of the DHL, is incorrect. It is in this region that the UVH suggests stronger feelings of strangeness compared with those evoked by neighboring less ambiguous human or humanlike stimuli. But the data show that greater feelings of strangeness are actually reported for the least human faces,

and that feelings of strangeness diminish with increasing human likeness of the facial morphs.

While there is no indication of an uncanny effect as described in the UVH, these data are based on group averaging of data. It is however possible that there are inter-individual differences in the relationship between familiarity and PD difficulty that are concealed by data averaging and that these differences might reveal an effect consistent with Mori's suggestion. In fact, the correlative data show a significant relationship between PD difficulty and feelings of familiarity, but the direction of this relationship is the opposite of that predicted in the UVH. Increasing PD difficulty is associated with more positive feelings of familiarity. Interestingly, this effect only applies for nonhuman and ambiguous faces. There was no significant relationship between PD difficulty and familiarity for human faces. Critically, this correlative effect was greatest for ambiguous faces. Taken together, the correlative data suggest, irrespective of the question of the causal direction, that the UVH' prediction is most likely to be wrong.

The reason for asymmetry in PD performance along the continua is not clear. One potential explanation draws on the suggestion that human observers preferentially code other members of the human in-group (e.g., our human exemplars) differently than members of a nonhuman out-group (e.g., our highly human-like avatars) (Cheetham et al., 2013; see the *other-race hypothesis*, Levin, 2000; *differential processing hypothesis*, Ostrom et al., 1993; *other-race effect*, Rhodes et al., 2006). This bias in coding means that individuals are tuned by categorization experience to detect subtle differences between other human individuals, thus facilitating face recognition among in-group members at the (individuating) exemplar level (see the *feature-selection hypothesis*, Levin, 2000). In contrast, individuals code information in the out-group that is more relevant for detection of out-group members, that is, information at the category level. At the category level, the best cognitive processing strategy for discriminating faces would be to code information indicating differences in human likeness along the DHL, thus enhancing discrimination of out-group members (i.e., our avatars). In contrast, a processing strategy that is more suited to face recognition of the individual human category exemplars than processing differences in human likeness along the DHL is more likely to result in poorer discrimination performance for human faces.

Face recognition among in-group members at the individuating level is more likely to rely on the use of configural information (Maurer et al., 2002), whereas there is evidence of less configural coding of out-group members (e.g., Rhodes et al., 1989; Fallshore and Schooler, 1995). Configural information relates to the individual arrangement of first- and second-order (e.g., nose-mouth distance) spatial relations among facial features (Rhodes, 1988). Configural processing is disrupted when faces are inverted instead of being presented upright (Diamond and Carey, 1986; Bartlett and Searcy, 1993; Rhodes et al., 1993; Rossion, 2009). If the asymmetry in PD between the avatar and human faces is attributable to a greater tendency to individuate human category exemplars than avatar category exemplars and a bias therefore toward greater configural processing of human exemplars, face inversion should reduce or eliminate the asymmetry. If on the other hand the asymmetry is not attributable to differences in configural processing, face inversion will have no impact on it. Experiment 3 was performed in order to test this.

EXPERIMENT 3

The ABX and forced choice categorization tasks were performed. The task requirements, instructions and stimulus presentation conditions for these tasks were identical to those described for the two preceding experiments, with one exception. The re-morphed stimuli that were presented in Experiment 2 were inverted by rotating them 180°.

PARTICIPANTS

A new sample of $N = 25$ volunteers (21 female, mean age 21 years; range 18–26 years) not involved in Experiments 1 or 2 participated in Experiment 3.

FORCED CHOICE CATEGORIZATION TASK: LOGISTIC FUNCTION, AND CATEGORY BOUNDARY

The parameter estimates derived from each logistic function model of each participant across continua were tested against zero in a one-sample *t*-test and showed, as in Experiments 1 and 2, a highly significant logistic component [$t_{(24)} = 22.29$, $p > 0.001$] (see **Figure 7**). Based on the parameter estimates β0 and β1, the mean category boundary value was $M = 6.7$. Across continua, the data show that the most ambiguous face morph M6 is closest to this boundary (**Figure 7**).

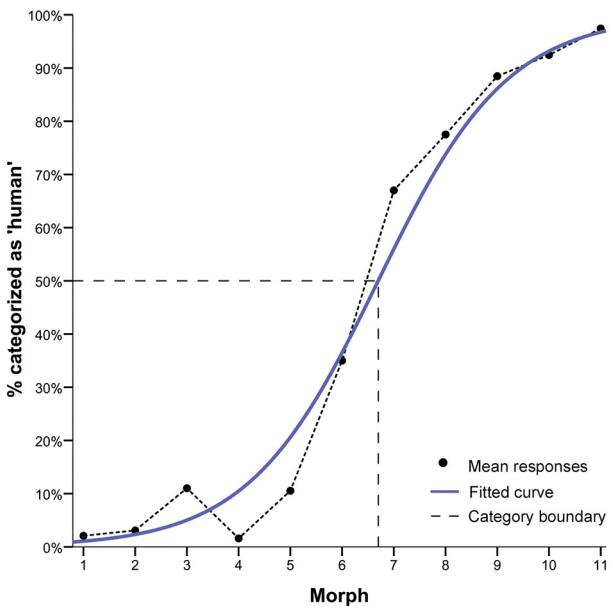


FIGURE 7 | Results of the forced choice classification task for inverted faces. Mean responses are depicted in terms of % of "human" responses, with the grand average across all continua (continuous blue line), the fitted logistic curve based on the grand mean (black line), and the category boundary (dashed gray line) to indicate the point of maximum uncertainty of 50% in categorisation judgements along the continua. Results indicate a step-like response function consistent with the presence of a category boundary.

For completeness, the other analyses for the forced choice categorization task conducted in Experiments 1 and 2 (i.e., categorization responses and RT) were repeated for Experiment 3. These produced the same pattern of results as Experiments 1 and 2 and are reported together with Figures in the *Supplemental information Experiment 3*.

RESULTS: ABX PERCEPTUAL DISCRIMINATION TASK

An independent two-sample *t*-test (Experiment 3 vs. Experiment 2) using mean d' of each participant across continua as dependent variable was conducted to compare PD performance in Experiments 2 and 3; these were compared because these experiments used the same re-morphed continua. This analysis showed that discrimination performance for each of the 9 morph pairs (i.e., M1-M3 to M9-M11) was not significantly different between Experiments 2 and 3. Levene's test of equality of variances indicated that the group variances for each of the 9 morph pairs could be treated as equal. For completeness, the same analysis was repeated to test for differences between Experiment 3 vs. Experiment 1. This showed a significant difference in discrimination between morph pairs M6-M8 [$t_{(72)} = 2.12, p > 0.038$] (note that M7 in Experiment 1 and M6 in Experiment 3 were the most ambiguous) and between the most human morph pairs M9-M11, [$t_{(72)} = 3.5, p > 0.001$]. There were no other differences (for the results of the three ABX experiments, showing all 9 morph pairs, see Supplemental Figure 7).

PD performance in Experiment 3 was then tested. Given that face morph position M6 was closest to the category boundary, differences in the ability to perceptually discriminate between pairs of morphs (M5–M7) straddling the *ambiguous* M6 compared with the ability to perceptually discriminate between unambiguous morphs within the *avatar* (M1-M3, M2-M4, M3-M5) and *human* (M7-M9, M8-M10, M9-M11) face categories were tested. This choice of morph pairs was based on the preceding data of the forced choice task, and ensured control for the physical morph distance between the ambiguous and unambiguous faces. The mean value of discrimination sensitivity was compared in a one-way RM-ANOVA with factor morph position (3 levels: avatar, ambiguous, human) using d' as dependent variable (see **Figure 8**). This analysis showed a significant effect for morph pair position, $F_{(2, 48)} = 11.18, p < 0.001$.

Tests of planned within-subject contrasts showed the same pattern of significant differences in PD as in Experiment 2. PD of avatar faces ($M = 1.44$, $SE = 0.13$) was significantly greater than that of faces within the human category ($M = 0.87$, $SE = 0.11$), $F_{(1, 48)} = 27.31, p > 0.001$. As in Experiment 2, discrimination sensitivity for ambiguous faces at the category boundary ($M = 1.32, SE = 0.15$) was not significantly greater than for faces within the avatar category, $F_{(1, 24)} = 1.03, p = 0.319$, but it was significantly greater than for faces within the human category, $F_{(1, 24)} = 9.5, p = 0.005$.

The data thus indicate that face inversion had no differential impact on the ability to discriminate between faces along the continua.

A one-way RM ANOVA with "morph position" (11 levels) and *c* as the dependent variable for response bias showed no significant differences for *c*.

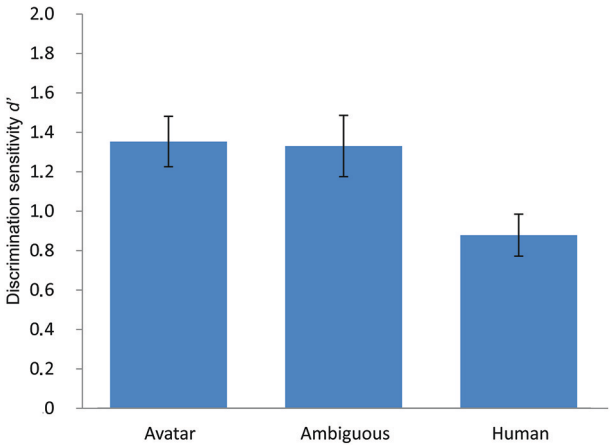


FIGURE 8 | Results of the ABX perceptual discrimination task for inverted faces. This figure depicts mean discrimination sensitivity d' in the ABX perceptual discrimination task for inverted unambiguous avatar and human and highly ambiguously faces. The data replicate those of experiments 1 and 2, showing the same asymmetry in perceptual discrimination performance along the dimension of human likeness. Face inversion had no impact on this, indicating that this asymmetry is not attributable to a differential processing strategy in which avatars are coded at a category and human faces at an individual level. The error bars indicate 1 *SE* ($N = 25$).

DISCUSSION

Experiment 3 explored the possibility that the asymmetry in PD reported in Experiments 1 and 2 might be attributable to a *differential processing bias*. This bias suggests that participants preferentially code human-category exemplars at the individual level and avatar-category exemplars at the category level. The data show that the inversion of faces had no impact on the asymmetry in PD, indicating that the asymmetry is not likely to be attributable to differences in configural coding and to a tendency to preferentially process human compared with avatar faces at an individual level.

GENERAL DISCUSSION

The UVH conceptualizes the DHL as a linear dimension of physical similarity space. This space is considered to span between points within a nonhuman category representing similar objects or characters of various degrees of human likeness and a single point representing the human category (**Figure 1**). The problem with this conceptualization and, more importantly, its faithful application in uncanny studies and theoretical considerations (e.g., Ramey, 2005; Tinwell et al., 2011) is that it implicitly assumes that this space does not vary within the human category. The assignment of physically different morphs to the human category in the forced choice categorization task clearly shows that this assumption is wrong (see also e.g., Looser and Wheatley, 2010; Cheetham et al., 2011; Yamada et al., 2013).

The advantage of considering the human end of the DHL is that it provides a basis of comparison for understanding how other objects and characters along the DHL are perceived and experienced. This approach is important for the present study. The UVH predicts enhanced negative affective experience as a

function of enhanced PD difficulty and suggests that this effect occurs at the point along the DHL at which categorization ambiguity is greatest. The data of the first and second experiments confirmed that there are differences in PD performance as a function of human likeness. But the pattern of differences in PD is very different than that implicitly assumed in the UVH. Firstly, and as expected on the basis of previous studies of CP, PD of faces at the category boundary is enhanced compared with PD of within-category human faces. Second, PD of within-category avatars is also enhanced compared with PD of within-category human faces. Together, these findings support the suggested asymmetry in PD along the DHL. In contrast to the UVH, they show that PD difficulty is greatest for human faces.

This finding of enhanced PD difficulty on the human side of the DHL's category boundary is reflected in the warped profile of psychological similarity space that is typically described for CP. This profile is characterized by attenuated PD performance for faces within the human category compared with enhanced PD performance for faces close to and at the category boundary (e.g., Livingston et al., 1998). In the present study, warping likely reflects the impact of perceptual and category learning processes over a person's history of everyday social interactive behavior with other members of the human category: All participants expressly reported no previous experience with our specific avatar parent faces, no previous experience with similarly humanlike faces (and robots), and no knowledge of previous experience with human likeness-related manipulations of perceptual features such as those applied along our morph continua. In contrast, they considered the human parent faces to be of the kind that they might typically encounter in normal everyday situations.

The impact of perceptual and category learning processes is that these likely lead to perceptual desensitization to within-category human features that are therefore perceived as more alike or *equivalent* and to enhanced perceptual sensitivity close to and at the category boundary to those stimulus features that facilitate assignment of category membership in everyday tasks (e.g., human vs. nonhuman). These features are therefore perceived as more *distinctive* (e.g., Lawrence, 1949; Gibson, 1991; Goldstone, 1994; Campbell et al., 1997; for an overview of *acquired distinctiveness* and *acquired equivalence*, see e.g., Goldstone, 1998). In contrast to the warped profile on the human side of the DHL's category boundary, there was no such difference in PD for unambiguous within-category avatar faces compared with the ambiguous faces at or closest to the category boundary. Considered in terms of the CP literature, participants thus appear to be perceptually desensitized to information that would facilitate visual discrimination of within-category human faces, while a corresponding desensitization is not apparent within the nonhuman category.

The present study did not aim to show that PD within the nonhuman category can change with perceptual and categorization experience. But stimulus exposure and explicit categorization training is known to evoke changes in discrimination sensitivity to a range of stimuli, from simple line drawings of unnatural entities to perceptually complex facial stimuli (e.g., Gibson, 1991; Hall, 1991; Schyns and Murphy, 1994; Goldstone, 1996; Levin, 1996, 2000; Livingston et al., 1998; Stevenage, 1998; Goldstone et al., 2003; Kikutani et al., 2008, 2010). If categorization training can modulate PD performance along the DHL, this might induce effects of acquired equivalence, acquired equivalence, or both, resulting therefore in a different profile of warping along the DHL than shown in the present study. Presumably, categorization training would primarily influence the cognitive representation of the avatar side of the DHL. Training could be based, for example, on familiarization with avatar faces so that individuals learn to discriminate between these in terms of their unique features (Bruyer et al., 2004; McGugin et al., 2011). Alternatively, the impact of experience might be examined in designers. Animators, video game designers, and roboticists concerned about the uncanny effect and the impact of their designs on subjective affect (e.g., Minato et al., 2006; Walters et al., 2008; MacDorman et al., 2009) regularly expose themselves to a range of humanlike faces and actively engage in carefully crafting perceptual features related to human likeness. Differences between novices and experts in processing perceptual information has been reported for other domains of expertise, ranging from the diagnosis of aberrant structures in x-rays to identification of gender in chickens (e.g., Burns and Ward, 1978; Biederman and Shiffrar, 1987; Myles-Worsley et al., 1988; Peron and Allen, 1988; Norman et al., 1992). This has yet to be examined in the present context.

In view of this asymmetry in PD performance, the third experiment examined whether avatars are preferentially coded at the category level and human faces at the exemplar level. This idea draws on findings relating to the *other-race affect* that show greater accuracy recognizing individual own- compared with other-race faces and show less configural coding of out-group members (e.g., Rhodes et al., 1989, 2006). The third experiment thus used inverted faces because inversion strongly influences efficient configural coding of spatial relations (e.g., nose-mouth distance) among facial features (Leder and Bruce, 2000), while its impact on processing the individual features is generally much weaker (e.g., Murray et al., 2000). The lack of an inversion effect in the present experiment suggests that PD performance along the DHL generally relies more on coding human likeness-specifying information of facial features such as the eyes, nose, and mouth and other features such as skin tone rather than on coding the spatial relationship among these features, even though coding configural information might enhance the accuracy of coding facial features (Tanaka and Farah, 1993).

The potential role of these facial features in the reported asymmetry in PD is therefore worth considering in terms of the *avatar-feature hypothesis* (Cheetham et al., 2013). This hypothesis initially related to categorization performance along the DHL. It suggests that participants preferentially detect perceptual information in nonhuman faces that is diagnostic of the nonhuman category. Assuming that it is cognitively less demanding to detect the presence of this diagnostic information in avatars rather than its absence in human faces, a categorization decision strategy based on "avatar vs. not avatar" instead of "avatar vs. human" would result in faster categorization decisions for avatars (see also *feature asymmetry*, Treisman and Gormican, 1988). Consistent with this, the forced choice categorization data of all three experiments show shorter categorization response latencies for avatar

compared with human faces, replicating the data of previous studies (Cheetham et al., 2013, 2011; see also Levin, 1996).

It is similarly possible that in the ABX tasks participants preferentially detected or found it easier to detect perceptual information that is diagnostic of human likeness specifically in the nonhuman faces of the DHL, thus facilitating the asymmetric effect in PD for these faces. The absence of an inversion effect in the ABX task indicates that this information is not likely to be relational (i.e., based on configural coding). Given that inversion effects are weaker for facial features like the eyes, nose and mouth and absent for facial properties like facial color (Leder and Carbon, 2006), it is conceivable that the participants coded and processed perceptual differences along the DHL on the basis of facial properties such as smoothed skin texture, color and shading. This does not exclude a role for feature-based processing, especially as processing for example the general luminance properties of faces can enhance processing of facial features (Sergent, 1986; Schyns and Oliva, 1994; Schyns and Gosselin, 2003). The question is why these properties should be easier to detect in the avatar faces. In view of the task context of processing novel avatars and everyday human faces, it is possible that perceptual information indicating the novelty of these facial properties renders this information more salient in the nonhuman faces of the DHL and that novelty therefore serves as a primitive perceptual feature that can facilitate PD within the avatar category (Levin, 2000). An alternative suggestion is that visual PD performance might be facilitated by the progressive reduction in perceptual complexity of the morphs with increasing distance from the human end of the continua independently of experience and perceptual strategy; the avatar parent faces have less human structural and textural detail than the human parent faces. Reduced humanlike complexity such as the reduced variance in shading of the smoothed skin texture might in itself provide a more easily detectable feature of these morphs that eases PD.

The UVH predicts that greater PD will evoke greater feelings of strangeness (i.e., feelings of less familiarity) at the point along the DHL at or near which ambiguity is greatest. The data of the second experiment suggest that this is wrong on two counts. Firstly, the analysis of familiarity ratings indicates that greater feelings of strangeness (i.e., feelings of less familiarity) are not reported for ambiguous faces. Instead, feelings of strangeness increased with increasing morph distance from the human end of the continua. This is consistent with the pattern reported in other studies in which comparably well-controlled morph continua and *ad hoc* measures of shinwakan such as measures of pleasantness have been used (e.g., Looser and Wheatley, 2010). These empirical data contradict the theoretical model of the UVH's uncanny valley effect presented by Moore (2012). Two drawbacks of that model is that it assumes *a priori* that the uncanny curve in Mori's graphical representation of the UVH is correct and it does not consider the potential impact of asymmetry in perceptual and categorization experience that is implicit in the UVH. The overall implication of the present familiarity data is that more humanlike stimuli simply evoke more positive affective experience and are preferred over less humanlike stimuli. The most straightforward explanation for this relates to the *mere-exposure* effect (Zajonc, 1968). This means that repeated exposure to human faces over a person's history of social interaction and the often more positive affective tone of interaction with particular in-groups results in more positive evaluations of other in-group members (e.g., Reis and Gable, 2003).

Second, the inter-individual differences approach adopted in the second experiment shows that there is indeed a significant relationship between familiarity and PD difficulty, but that the direction of this relationship is the opposite of that predicted in the UVH. Increasing PD difficulty is associated with more positive feelings of familiarity. The effect was evident for nonhuman and ambiguous faces, whereas there was no significant relationship between PD and feelings of familiarity for human faces. This correlative effect was greatest for ambiguous faces, indicating that, irrespective of the causal relationship between PD and feelings of familiarity, the UVH' prediction is most likely to be incorrect. It should be noted that the UVH does not suggest how DP difficulty and its affective dimension, shinwakan, should be operationalised. This issue has hampered uncanny-related research from the outset. But the approach taken in the present study to testing the relationship between DP and affective experience (as described in the UVH) was straightforward and produces strong effects, indicating that further examination of this relationship might be fruitful.

Why greater PD difficulty should correlate with more positive self assessment of affect is not clear. A popular account of the uncanny effect is based on the *Hedonic Fluency Model* (Winkielman et al., 2003; see Yamada et al., 2013). This suggests that negative evaluations of novel or unfamiliar stimuli relate to cognitive difficulty extracting information needed for rapid and efficient processing. This makes sense if the UVH' prediction for PD is assumed to be correct. But the present data suggest that this prediction is incorrect. The present PD data do however fit better with an alternative model of processing fluency, the *Fluency Amplification Model* (Albrecht and Carbon, 2014). This model states that processing fluency enhances the affective reaction that the stimulus already evokes. Assuming for example that the valence of a given stimulus is initially experienced as comparatively negative, individuals who experience greater fluency (in our case, lesser difficulty in PD) will experience the negative stimulus as even more negative. By the same token, greater PD difficulty would correlate with less negative ratings. While this interpretation is consistent with the present correlative data, further investigation of this finding and of the role of interindividual differences in state affect is needed.

The ABX PD task is useful for testing naive participants because it requires no description of the specific physical dimensions along which the stimuli vary and participants do not need to know the category labels. One explanation for CP effects suggests a role for the presence of category labels (Roberson and Davidoff, 2000; Pilling et al., 2003; Kikutani et al., 2008). This is because within-category stimuli differ only at the exemplar level, while cross-category stimuli differ both at the exemplar and category levels. If exemplar-level information and category-level information are processed in parallel so that the category boundary can be represented in naive participants after initial learning (Marsolek, 2004), category-level processing might encourage the use of labeling and the emergence of CP effects. This effect might

be even stronger in a task with a strong memory component such as the ABX task (i.e., the test stimulus must be compared with the stored representation of the target stimuli). Considering the asymmetry in discrimination performance around the category boundary, a labeling effect is unlikely, unless labeling affected the human side of the category boundary only. It has however been argued that any impact of category labeling would be reflected in specific *within-category discrimination asymmetries* (Hanley and Roberson, 2011). There are no such asymmetries within the human or nonhuman categories.

In summary, the data of the three experiments reject the implicit assumption underlying the UVH' key prediction. The data show lesser PD difficulty for categorically ambiguous faces and for unambiguous avatar faces and, notably, greater PD difficulty for unambiguous human faces. The data indicate that this asymmetry in PD difficulty cannot be attributed to differences between human and nonhuman faces in configural coding. It is likely that perceptual differences along the DHL are generally processed on the basis of human likeness-related manipulations of facial properties such as skin texture, color and shading. Ratings of familiarity show that faces associated with greatest category ambiguity do not show an uncanny-like effect. Negatively valenced ratings increased across the tested stimulus conditions with increasing distance from the human end of the continua. An interindividual differences approach revealed that greater PD difficulty is associated with more positively rather than negatively valenced experience. This challenges the key idea behind the UVH. This effect is strongest for ambiguous faces, suggesting that this effect is more consistent with the metaphor "happy valley" and, correspondingly, the fluency amplification effect. These findings for our static faces thus indicate that both the assumed distribution of PD difficulty along the DHL and the predicted relationship between PD difficulty and affective experience (as "conceptualized" in the UVH) are very likely wrong.

Clearly, it is not possible to confirm or refute the vaguely formulated non-scientific UVH in its current form. Our approach has been to augment the notions underlying the UVH with the necessary assumptions needed to render the essential features of the hypothesis testable. While we find no evidence in favor of these notions, our findings do not exclude the possibility that alternative experimental paradigms and other methodologies might show effects consistent with the underlying idea of the UVH. It should be noted that only male face stimuli were presented. The choice of stimuli for this study was not guided by the well-known depiction of Mori's hypothesis in **Figure 1** because we sought to ensure that perceptual discriminative, categorization and familiarity judgments would not be confounded by factors other than the manipulation of human likeness (for a discussion of confounds, see Cheetham and Jancke, 2013). This study presented stimuli similar to those used in preceding studies (Cheetham et al., 2011, 2013), which, given the absence of comparable paradigms in the investigation of the DHL, has provided an effective means to developing insight and a basis for further uncanny-related study. But an important element of further study would be to examine whether these findings generalize to other static stimuli. Whether these findings might apply to dynamic nonhuman characters (e.g., Saygin et al., 2012; Burleigh et al., 2013; Urgen et al., 2013) and to such characters in human interaction (e.g., Cheetham et al., 2009) is open to further investigation.

ACKNOWLEDGMENTS

This work was supported by the European Union FET Integrated Project PRESENCCIA (Contract number 27731).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpsyg.2014.01219/abstract

REFERENCES

Albrecht, S., and Carbon, C. C. (2014). The Fluency Amplification Model: fluent stimuli show more intense but not evidently more positive evaluations. *Acta Psychol.* 148, 195–203. doi: 10.1016/j.actpsy.2014.02.002

Annett, M. A., classification of hand preference by association analysis. (1970). *Br. J. Psychol.* 61, 303–321. doi: 10.1111/j.2044-8295.1970.tb01248.x

Bartlett, J. C., and Searcy, J. (1993). Inversion and configuration of faces. *Cogn. Psychol.* 25, 281–316. doi: 10.1006/cogp.1993.1007

Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. (2007). "Is the uncanny valley an uncanny cliff?" in *Proceedings of the 16th IEEE, RO-MAN.* (Jeju), 368–373.

Beale, J. M., and Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition* 57, 217–219. doi: 10.1016/0010-0277(95)00669-X

Best, C. T., Morrongiello, B., and Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Percept. Psychophys.* 29, 191–211. doi: 10.3758/BF03207286

Biederman, I., and Shiffrar, M. (1987). Sexing day-old chicks: a case study and expert systems analysis of a difficult perceptual learning task. *J. Exp. Psychol. Learn. Mem. Cogn.* 13, 640–645. doi: 10.1037/0278-7393.13.4.640

Bruyer, R., Leclere, S., and Quinet, P. (2004). Ethnic categorisation of faces is not independent of face identity. *Perception* 33, 169–179. doi: 10.1068/p5094

Bülthoff, I., and Newell, F. N. (2000). Investigating categorical perception of gender with 3-D morphs of familiar faces. *Perception* 29, 57. doi: 10.1068/p2990

Burleigh, T. J., Schoenherr, J. R., and 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. *Comput. Hum. Behav.* 29, 759–771. doi: 10.1016/j.chb.2012.11.021

Burns, E. M., and Ward, W. D. (1978). Categorical perception– phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals. *J. Acoust. Soc. Am.* 63, 456–468. doi: 10.1121/1.381737

Campbell, R., Pascalis, O., Coleman, M., Wallace, S. B., and Benson, P. J. (1997). Are faces of different species perceived categorically by human observers? *Proc. Biol. Sci.* 264, 1429–1434. doi: 10.1098/rspb.1997.0199

Cheetham, M., and Jancke, L. (2013). Perceptual and category processing of the Uncanny Valley Hypothesis' dimension of human likeness: some methodological issues. *J. Vis. Exp.* e4375. doi: 10.3791/4375

Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., and Jancke, L. (2013). Category processing and the human likeness dimension of the Uncanny Valley Hypothesis: eye-tracking data. *Front. Psychol.* 4:108. doi: 10.3389/fpsyg.2013.00108

Cheetham, M., Pedroni, A. F., Antley, A., Slater, M., and Jäncke, L. (2009). Virtual milgram: empathic concern or personal distress? Evidence from functional MRI and dispositional measures. *Front. Hum. Neurosci.* 3, 29. doi: 10.3389/neuro.09.029.2009

Cheetham, M., Suter, P., and Jancke, L. (2011). The human likeness dimension of the "uncanny valley hypothesis": behavioral and functional MRI findings. *Front. Hum. Neurosci.* 5:126. doi: 10.3389/fnhum.2011.00126

Diamond, R., and Carey, S. (1977). Development changes in the representation of faces. *J. Exp. Child Psychol.* 23, 1–22. doi: 10.1016/0022-0965(77)90069-8

Diamond, R., and Carey, S. (1986). Why faces are and are not special: an effect of expertise. *J. Exp. Psychol. Gen.* 115, 107–117. doi: 10.1037/0096-3445.115.2.107

Dill, V., Flach, L., Hocevar, R., Lykawka, C., Musse, S., and Pinho, M. (2012). "Evaluation of the uncanny valley in CG characters," in *Intelligent Virtual Agents Lecture Notes in Computer Science*, eds N. Yukiko, M. Neff, A. Paiva, and M. Walker (Berlin: Springer), 511–513.

Ectoff, N. L., and Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition* 44, 227–240. doi: 10.1016/0010-0277(92)90002-Y

Fallshore, M., and Schooler, J. W. (1995). The verbal vulnerability ofperceptual expertise. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 1608–1623. doi: 10.1037/0278-7393.21.6.1608

Francis, A. L., and Ciocca, V. (2003). Stimulus presentation order and the perception of lexical tones in Cantonese. *J. Acoust. Soc. Am.* 114, 1611–1621. doi: 10.1121/1.1603231

Gibson, E. J. (1991). *An Odyssey in Learning and Perception.* Cambridge: MIT Press.

Goldstone, R. (1994). Influences of categorization of perceptual discrimination. *J. Exp. Psychol. Gen.* 123, 178–200. doi: 10.1037/0096-3445.123.2.178

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Mem. Cogn.* 24, 608–628. doi: 10.3758/BF03201087

Goldstone, R. L. (1998). Perceptual learning. *Annu. Rev. Psychol.* 49, 585–612. doi: 10.1146/annurev.psych.49.1.585

Goldstone, R. L., and Hendrickson, A. T. (2010). Categorical perception. *Interdiscipl. Rev. Cogn. Sci.* 1, 65–78. doi: 10.1002/wcs.26

Goldstone, R. L., Steyvers, M., and Rogosky, B. J. (2003). Conceptual interrelatedness and caricatures. *Mem. Cognit.* 31, 169–180. doi: 10.3758/BF03194377

Green, R. D., MacDorman, K. F., Ho, C.-C., and Vasudevan, S. K. (2008). Sensitivity to the proportions of faces that vary in human likeness. *Comput. Hum. Behav.* 24, 2456–2474. doi: 10.1016/j.chb.2008.02.019

Hall, G. (1991). *Perceptual and Associative Learning.* Oxford: Clarendon Press.

Hanley, J. R., and Roberson, D. (2011). Categorical perception effects reflect differences in typicality on within-category trials. *Psychon. Bull. Rev.* 18, 355–363. doi: 10.3758/s13423-010-0043-z

Hanson, D. (2006). "Exploring the aesthetic range for humanoid robots," in *Proceedings of the ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science* (Vancouver), 39–42.

Harnad, S. (1987). *Categorical Perception*. Cambridge: Cambridge University Press.

Hautus, M. J., and Meng, X. (2001). Decision strategies in the ABX (matching-to-sample) psychophysical task. *Percept. Psychophysiol.* 64, 89–106. doi: 10.3758/BF03194559

Ho, C.-C., and MacDorman, K. (2010). Revisiting the uncanny valley theory: developing and validating an alternative to the Godspeed indices. *Comput. Hum. Behav.* 26, 1508–1518. doi: 10.1016/j.chb.2010.05.015

Kikutani, M., Roberson, D., and Hanley, J. R. (2008). What's in the name? Categorical perception of unfamiliar faces can occur through labeling. *Psychon. Bull. Rev.* 15, 787–794. doi: 10.3758/PBR.15.4.787

Kuhl, P. K. (1991). Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories monkeys do not. *Percept. Psychophysiol.* 50, 93–107. doi: 10.3758/BF03212211

Kikutani, M., Roberson, D., and Hanley, J. R. (2010). Categorical perception for unfamiliar faces: effect of covert and overt face learning. *Psychol. Sci.* 21, 865–872. doi: 10.1177/0956797610371964

Lawrence, D. H. (1949). Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. Exp. Psychol.* 39, 770–784. doi: 10.1037/h0058097

Leder, H., and Bruce, V. (2000). When inverted faces are recognized: the role of configural information in face recognition. *Q. J. Exp. Psychol. A* 53, 513–536. doi: 10.1080/713755889

Leder, H., and Carbon, C. C. (2006). Face-specific configural processing of relational information. *Br. J. Psychol.* 97, 19–29. doi: 10.1348/000712605X 54794

Levin, D. T. (1996). Classifying faces by race: the structure of face categories. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 1364–1382. doi: 10.1037/0278-7393.22.6.1364

Levin, D. T. (2000). Race as a visual feature: using visual search and perceptual discrimination tasks to understand face categories and the cross race recognition deficit. *J. Exp. Psychol. Gen.* 129, 559–574. doi: 10.1037/0096-3445.129.4.559

Levin, D. T., and Beale, J. M. (2000). Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Percept. Psychophysiol.* 62, 386–401. doi: 10.3758/BF03205558

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* 54, 358–368. doi: 10.1037/h0044417

Livingston, K. R., Andrews, J. K., and Harnad, S. (1998). Categorical perception eVects induced by category learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 732–753. doi: 10.1037/0278-7393.24.3.732

Looser, C. E., and Wheatley, T. (2010). The tipping point of animacy: how, when, and where we perceive life in a face. *Psychol. Sci.* 21, 1854–1862. doi: 10.1177/0956797610388044

MacDorman, K. (2005). "Androids as an experimental apparatus: why is there an uncanny valley and can we exploit it?" in *Toward Social Mechanisms of Android Science: A CogSci 2005 Workshop* (Stresa), 106–118.

MacDorman, K. F. (2006). "Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: an exploration of the uncanny valley [electronic resource]," in *ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science* (Vancouver).

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

MacDorman, K. F., Srinivas, P., and Patel, H. (2013). The uncanny valley does not interfere with level 1 visual perspective taking. *Comput. Human Behav.* 29, 1671–1685. doi: 10.1016/j.chb.2013.01.051

MacDorman, K., Green, R., Ho, C.-C., and Koch, C. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Comput. Human Behav.* 25, 695–710. doi: 10.1016/j.chb.2008.12.026

Macmillan, N. A., and Creelman, C. D. (2005). *Detection Theory: A User's Guide*. Mahwah, NJ: Lawrence Erlbaum Associates.

Marsolek, C. (2004). Abstractionist versus exemplar-based theories of visual word priming: a subsystems resolution. *Q. J. Exp. Psychol.* 57A, 1233–1259. doi: 10.1080/02724980343000747

Maurer, D., Le Grand, R., and Mondloch, C. J. (2002). The many faces of configural processing. *Trends Cogn. Sci.* 6, 255–260. doi: 10.1016/S1364-6613(02)01903-4

McGugin, R. W., Tanaka, J. W., Lebrecht, S., Tarr, M. J., and Gauthier, I. (2011). Race-specific perceptual discrimination improvement following short individuation training with faces. *Cogn. Sci.* 35, 330–347. doi: 10.1111/j.1551-6709.2010.01148.x

Minato, T., Shimada, M., Itakura, S., Lee, K., and Ishiguro, H. (2006). Evaluating human likeness of an android by comparing gaze behaviors elicited. *Adv. Robot.* 20, 1147–1163. doi: 10.1163/156855306778522505

Moore, R. K. (2012). A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena. *Sci. Rep.* 2:864. doi: 10.1038/srep00864

Mori, M. (1970). "Bukimi no tani [The uncanny valley]. Energy, *7*(4) 33-35. (Translated by Karl F. MacDorman and Takashi Minato in 2005 within Appendix B for the paper Androids as an Experimental Apparatus: Why is there an uncanny and can we exploit it?" in *Proceedings of the CogSci-2005 Workshop: Toward Social Mechanisms of Android Science* (Italy), 106–118.

Mori, M. (2012). The uncanny valley (K. F. MacDorman & Norri Kageki, Trans.). *IEEE Robot. Autom.* 19, 98–100. doi: 10.1109/MRA.2012.2192811

Murray, J. E., Yong, E., and Rhodes, G. (2000). Revisiting the perception of upside-down faces. *Psychol. Sci.* 11, 498–502. doi: 10.1111/1467-9280.00294

Myles-Worsley, M., Johnston, W. A., and Simons, M. A. (1988). The influence of expertise on X-ray image processing. *J. Exp. Psychol. Learn. Mem. Cogn.* 14, 553–557. doi: 10.1037/0278-7393.14.3.553

Newell, F. N., and Bulthoff, H. H. (2002). Cognition Categorical perception of familiar objects. *Cognition* 85, 113–143. doi: 10.1016/S0010-0277(02)00104-X

Norman, D. A., Coblentz, C. L., Brooks, L. R., and Babcook, C. J. (1992). Expertise in visual diagnosis: a review of the literature. *Acad. Med.* 67, S78–S83. doi: 10.1097/00001888-199210000-00045

Ostrom, T. M., Carpenter, S. L., Sedikides, C., and Li, F. (1993). Differential processing of in-group and out-group information. *J. Pers. Soc. Psychol.* 64, 21–34. doi: 10.1037/0022-3514.64.1.21

Schyns, P. G., and Gosselin, F. (2003). "Diagnostic use of scale information fo rcomponential and holistic recognition," in *Perception of Faces, Objects, and Scenes. Analytic and Holistic Processes*, eds M. A. Peterson and G. Rhodes (Oxford: Oxford University Press), 120–145.

Peron, R. M., and Allen, G. L. (1988). Attempts to train novices for beer flavor discrimination: a matter of taste. *J. Gen. Psychol.* 115, 403–418. doi: 10.1080/00221309.1988.9710577

Pierce, J. R., and Gilbert, E. N. (1958). On AX and ABX limens. *J. Acoust. Soc. Am.* 30, 593–595. doi: 10.1121/1.1909700

Pilling, M., Wiggett, A., Özgen, E., and Davies, I. R. L. (2003). Is color "categorical perception" really perceptual? *Mem. Cognit.* 31, 538–551. doi: 10.3758/BF03196095

Prins, N., and Kingdom, F. A. A. (2009). *Palamedes: Matlab Routines for Analyzing Psychophysical Data*. Available Online at: http://www.palamedestoolbox.org

Ramey, C. H. (2005). "The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots," in *Paper Presented at the Proceedings of Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots*, (Tsukuba).

Reis, H. T., and Gable, S. L. (2003). "Toward a positive psychology of relationships," in *Flourishing: The Positive Person and the Good Life,* eds C. L. Keyes and J. Haidt (Washington, DC: American Psychological Association), 129–159.

Rhodes, G. (1988). Looking at faces: first-order and second-order features as determinants of facial appearance. *Perception* 17, 43–63. doi: 10.1068/p170043

Rhodes, G., Brake, S., and Atkinson, A. P. (1993). What's lost in inverted faces? *Cognition* 47, 25–57. doi: 10.1016/0010-0277(93)90061-Y

Rhodes, G., Hayward, W. G., and Winkler, C. (2006). Expert face coding: configural and component coding of own-race and other-race faces. *Psychon. Bull. Rev.* 13, 499–505. doi: 10.3758/BF03193876

Rhodes, G., Tan, S., Brake, S., and Taylor, K. (1989). Expertise and configural coding in face recognition. *Br. J. Psychol.* 80, 313–331. doi: 10.1111/j.2044-8295.1989.tb02323.x

Roberson, D., and Davidoff, J. (2000). The categorical perception of colors and facial expressions: the effect of verbal interference. *Mem. Cognit.* 28, 977–986. doi: 10.3758/BF03209345

Rossion, B. (2009). Distinguishing the cause and the consequence of face inversion: the perceptual field hypothesis. *Acta Psychol.* 132, 300–312. doi: 10.1016/j.actpsy.2009.08.002

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., and Frith, C. F. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* 7, 413–422. doi: 10.1093/scan/nsr025

Schultz, J., and Pilz, K. S. (2009). Natural facial motion enhances the cortical responses to faces. *Exp. Brain Res.* 194, 465–475. doi: 10.1007/s00221-009-1721-9

Schyns, P. G., and Murphy, G. L. (1994). The ontogeny of part representation in object concepts. *Psychol. Learn. Motiv.* 31, 301–349.

Schyns, P. G., and Oliva, A. (1994). From blobs to boundary edges: evidence for time and spatial scale depen dent scene recognition. *Psychol. Sci.* 5, 195–200. doi: 10.1111/j.1467-9280.1994.tb00500.x

Sergent, J. (1986). "Microgenesis of face perception," in *Aspects of Face Processing*, eds H. D. Ellis, M. A. Jeeves, F. Newcombe, and A. M. Young (Dordrecht: Martinus Nijhoff), 17–33.

Seyama, J. I., and Nagayama, R. S. (2007). The uncanny valley: effect of realism on the impression of artificial human faces. *Presence Teleoperat. Virtual Environ.* 16, 337–351. doi: 10.1162/pres.16.4.337

Sigala, R., Logothetis, N. K., and Rainer, G. (2011). Own-species bias in the representations of monkey and human face categories in the primate temporal lobe. *J. Neurophysiol.* 105, 2740–2752. doi: 10.1152/jn.00882.2010

Stevenage, S. V. (1998). Which twin are you? A demonstration of induced categorical perception of identical twin faces. *Br. J. Psychol.* 89, 39–57. doi: 10.1111/j.2044-8295.1998.tb02672.x

Tanaka, J. W. (2001). The entry point of face recognition: evidence for face expertise. *J. Exp. Psychol. Gen.* 130, 534–543. doi: 10.1037/0096-3445.130.3.534

Tanaka, J. W., and Farah, M. J. (1993). Parts and wholes in face recognition. *Q. J. Exp. Psychol.* 46A, 225–245. doi: 10.1080/14640749308401045

Tinwell, A., Grimshaw, M., and Williams, A. (2011). The Uncanny Wall. *Int. J. Arts Technol.* 4, 326–341. doi: 10.1504/IJART.2011.041485

Treisman, A., and Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychol. Rev.* 95, 15–48. doi: 10.1037/0033-295X.95.1.15

Urgen, B. A., Plank, M., Ishiguro, H., Poizner, H., and Saygin, A. P. (2013). EEG Theta and Mu oscillations during perception of human and robot actions. *Front. Neurorobot.* 7:19. doi: 10.3389/fnbot.2013.00019

Van Selst, M., and Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *Q. J. Exp. Psychol.* 47A, 631–650. doi: 10.1080/14640749408401131

Walters, M. L., Syrdal, D. S., Dautenhaun, K., Boekhorst, R., and Koay, K. L. (2008). Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Auton. Robots* 24, 159–178. doi: 10.1007/s10514-007-9058-3

Winkielman, P., Schwarz, N., Fazendeiro, T., and Reber, R. (2003). "The hedonic marking of processing fluency: implications for evaluative judgment," in *The Psychology of Evaluation: Affective Processes in Cognition and Emotion*, ed J. M. K. C. Klauer (Mahwah, NJ: Lawrence Erlbaum), 189–217.

Yamada, Y., Kawabe, T., and Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *Jpn. Psychol. Res.* 55, 20–32. doi: 10.1111/j.1468-5884.2012.00538.x

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *J. Pers. Soc. Psychol. Monogr.* 9(2 Pt 2), 1–27. doi: 10.1037/h0025848

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 10 July 2013; accepted: 08 October 2014; published online: 19 November 2014.

*Citation: Cheetham M, Suter P and Jancke L (2014) Perceptual discrimination difficulty and familiarity in the Uncanny Valley: more like a "Happy Valley". Front. Psychol. **5**:1219. doi: 10.3389/fpsyg.2014.01219*

This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 21 January 2015
doi: 10.3389/fpsyg.2014.01488

A reappraisal of the uncanny valley: categorical perception or frequency-based sensitization?

***Tyler J. Burleigh*[1]* and *Jordan R. Schoenherr*[2]**

[1] *Department of Psychology, University of Guelph, Guelph, ON, Canada*
[2] *Department of Psychology, Carleton University, Ottawa, ON, Canada*

Edited by:
Marcus Cheetham, University of Zürich, Switzerland

Reviewed by:
Rosemarie Velik, Carinthian Tech Research, Austria
Anthony Paul Atkinson, Durham University, UK

****Correspondence:***
Tyler J. Burleigh, Department of Psychology, University of Guelph, 50 Stone Road East, Guelph, ON N1G 2W1, Canada
e-mail: tburleig@uoguelph.ca

The uncanny valley (UCV) hypothesis describes a non-linear relationship between perceived human-likeness and affective response. The "uncanny valley" refers to an intermediate level of human-likeness that is associated with strong negative affect. Recent studies have suggested that the uncanny valley might result from the categorical perception of human-like stimuli during identification. When presented with stimuli sharing human-like traits, participants attempt to segment the continuum in "human" and "non-human" categories. Due to the ambiguity of stimuli located at a category boundary, categorization difficulty gives rise to a strong, negative affective response. Importantly, researchers who have studied the UCV in terms of categorical perception have focused on categorization responses rather than affective ratings. In the present study, we examined whether the negative affect associated with the UCV might be explained in terms of an individual's degree of exposure to stimuli. In two experiments, we tested a frequency-based model against a categorical perception model using a category-learning paradigm. We manipulated the frequency of exemplars that were presented to participants from two categories during a training phase. We then examined categorization and affective responses functions, as well as the relationship between categorization and affective responses. Supporting previous findings, categorization responses suggested that participants acquired novel category structures that reflected a category boundary. These category structures appeared to influence affective ratings of eeriness. Crucially, participants' ratings of eeriness were additionally affected by exemplar frequency. Taken together, these findings suggest that the UCV is determined by both categorical properties as well as the frequency of individual exemplars retained in memory.

Keywords: uncanny valley, categorical perception, category learning, categorization, exemplar theory, exemplar-based, frequency-based, affect

INTRODUCTION

Categorization is a critical determinant of human survival. In the absence of categories, humans would be required to learn whether each stimulus that we encountered was desirable or noxious as well as whether the conspecifics that we encountered were kin or competitors. The variability in cross-cultural folktaxonomies (Medin and Atran, 1999), color classification (Regier and Kay, 2009), and speech perception (Pisoni et al., 1982) demonstrates that while humans might have prepotent responses to ranges of stimuli, many of these distinctions can be modified or must be learned. When available within the classification system of society, these categories can be associated with strong, negative affect responses (Schoenherr and Burleigh, 2014). Thus, categories both reflect and determine one's experience of the world.

Group membership and identity form an especially relevant class of categories for humans (for a review, see Fiske and Taylor, 2013). In the social context, repeated exposure to individuals within a group can increase affiliation and conformity (for review, see Bond and Smith, 1996) among group members while also leading to negative affective responses toward out-group members (for review, see Cialdini and Goldstein, 2004). This suggests the possibility that mixing features that have strong associations with members of contrasting categories will either lead to a reduction in positive affect or an increase in negative affect (Burleigh et al., 2013). In contrast to categorical perception, sub-categorical properties such as exposure to individual exemplars has long been considered an important determinant of affective responses (e.g., Fechner, 1876; Maslow, 1937; Zajonc, 1968). The present study considers how the comparatively low frequency of exposure to stimuli selected from a region of a continuum can lead to negative affective responses. We examine this in the context of negative affective responses to stimuli containing features from contrasting categories.

In the context of human factors, Mori's (1970) Uncanny Valley Hypothesis (UVH) suggests that human-like objects in our environment might come to be associated with negative affect if they possess a certain degree of human-likeness. Recently, a number of authors have suggested potential explanations of the UVH that are either explicitly or implicitly based on categorical perception (Cheetham et al., 2011, 2013; Moore, 2012; Burleigh et al., 2013;

Yamada et al., 2013; Ferrey et al., submitted). While these studies have made important theoretical contributions, the implications of different learned category representations on the UCV phenomenon have not been directly tested. In the present study, we sought to address this gap by using a category-learning paradigm in which groups of participants received different sets of training stimuli consisting of computer-generated creatures. We examine participants responses to creatures following training, specify the conditions in which affective minima associated with the UCV would be observed, the properties of category learning that would determine the location of affective minima, and what underlying representation of category structures would best fit the response patterns that we observed.

THE UNCANNY VALLEY

While the essential phenomenon of the uncanny valley has a number of cultural antecedents (Schoenherr and Burleigh, 2014), the Uncanny Valley Hypothesis (UVH) remains underspecified. Mori (1970) initially proposed that human-like stimuli can elicit positive or negative feelings depending on their degree of similarity to humans. In contrast to the linear relationship between familiarity and positive affect for human and human-like faces (see Experiment 1, Burleigh et al., 2013), the UVH predicts a non-linear relationship. Mori's account assumes that as stimuli become defined by more human-like features, they will elicit greater positive affect. But importantly, his account also assumes that there is a critical region of intermediate human-likeness where a sharp decrease in positive feelings are observed. As illustrated in **Figure 1**, the proposed relationship resembles a cubic function, and the global minimum is referred to as the "valley."

It is important to distinguish between the classic and generalized forms of the UVH. The classic account of the UVH provided by Mori (1970) is defined as a non-linear function. In his account, the x-axis of this function is defined as human-likeness, and it is anchored by a non-human or minimally human-like entity at one end (e.g., a robot) and a real human at the other end. One reason to question this account is that it was informed by anecdotal evidence in the context of the human-like design of machines and other artifacts. The basic premise that negative affect could be a consequence of mismatch between features associated with contrasting classes would explain many cross-cultural phenomena (Schoenherr and Burleigh, 2014). In contrast to this, the UVH could be taken as assuming that the non-linear response function observed with human-likeness is a special case of more general cognitive and affective processes associated with stimulus frequency and categorical perception. Thus, it follows that similar non-linear phenomena should be observed in response to perceptual continua that represent non-human anchors with similar properties.

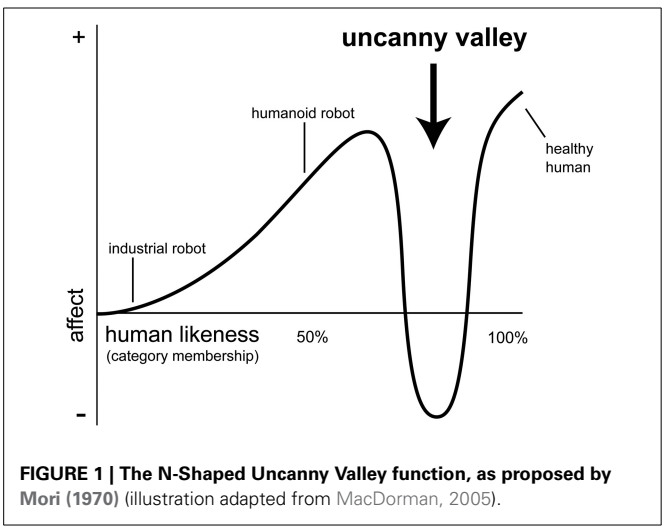


FIGURE 1 | The N-Shaped Uncanny Valley function, as proposed by Mori (1970) (illustration adapted from MacDorman, 2005).

Evidence that has been interpreted as supporting the classic UVH has been obtained from studies using a variety of stimuli selected from a number of ontological categories. A majority of these studies have observed affective functions that are consistent with the UVH when using stimuli representing computer-generated morph sequences of human and non-human entities, including non-human animals, robots, and anthropomorphic dolls (MacDorman and Ishiguro, 2006; Seyama and Nagayama, 2007; Burleigh et al., 2013; Ferrey et al., submitted). Many studies have also observed the affective function in response to images of existing artifacts that vary in human-likeness, such as androids, videogame characters, and prosthetic hands (Bartneck et al., 2007; Schneider et al., 2007; Poliakoff et al., 2013); however, it is worth noting that several studies have not found support for the classic account of the UVH (MacDorman et al., 2009; Cheetham et al., 2014). Across those studies which have found support, a general observation is that affective response is positively correlated with human-likeness, except at an intermediate level of human-likeness where there is a maximum of negative affect.

Few studies have examined the possibility that perceptual continua representing non-human entities could produce UCV phenomena. To the best of our knowledge, only two studies have examined this possibility. In Yamada et al. (2013, Experiment 2) morph sequences were generated that represented transitions between cartoon, stuffed, and real dogs. In Ferrey et al. (submitted, Experiment 1), bistable morph sequences were used that represented transitions between various non-human animals (e.g., between a duck and an elephant). In each of these studies, regions of maximal negative affect were found at intermediate levels of the perceptual continua, which is consistent with the generalized account of the UVH (see, Burleigh et al., 2013, Experiment 2). Between general formulations of the UVH and empirical support for UCV-like phenomena, greater theoretical consideration of the affective and cognitive processes is required to define the conditions under which the UCV will be observed as well as to differentiate it from related phenomena.

EXPLANATIONS FOR THE UNCANNY VALLEY PHENOMENON

Although the UVH provides a description of the non-linear response function, it does not explain why this function occurs, nor does it specify the mechanisms that are responsible. A common explanation is that the negative affect associated with uncanny stimuli might be a consequence of biological adaptations for threat avoidance behaviors (e.g., MacDorman et al., 2009). Stimuli within the valley might be convincing depictions

of humans while falling short of a satisfactory level of human-likeness due to imperfections. These imperfections might cause them to be seen as "humans with disease" which triggers an aversive response (MacDorman and Ishiguro, 2006; MacDorman et al., 2009; Burleigh et al., 2013, Experiment 2). There is some evidence supporting this account. For example, Ho et al. (2008) observed that disgust could explain a significant portion of the variance in eeriness ratings. Furthermore, Steckenfinger and Ghazanfar (2009) observed that macaque monkeys displayed an aversion (as measured with looking times) to images of conspecifics that were of intermediate realism, which suggests that there might be an evolutionary basis to the phenomenon.

From this account, it might be reasonable to assume that the UCV phenomenon is specific to observers viewing images of conspecifics—an assumption that would be consistent with the classic UVH. Given that the spread of communicable diseases depends on the genetic similarity between the observer and the stimulus entity, it is possible that a species could have evolved mechanisms that allow them to respond to pathogen cues in conspecifics, but not heterospecifics. Communicable diseases, however, are not the only source of contamination that members of a species have had to contend with in their environments. As Rozin and Fallon (1987) point out, disgust is also a food-related emotion, which serves to prevent the oral incorporation of contaminated substances. As Schoenherr and Burleigh (2014) discuss, food substances that share features from two categories have been associated with aversive responses, such as food taboos (e.g., some refer to a certain transgenic tomato as a "Frankenfood" because it incorporates genes from a winter flounder). This suggests that the UCV phenomenon might not be specific to observers viewing images of conspecifics, but that it might also occur more generally in response to the categorical ambiguity of certain types of stimuli. Even if these accounts are correct, general learning mechanisms would also allow for the adjustment of diagnostic features of disease as well as inclusion and exclusion of categories associated with disease as a result of an individual's experiences with their environment.

Another theory that accounts for threat avoidance behavior is based on the premise that appearances provide information that allows individuals to predict behavior, and thus to anticipate potential threats in their environment. Some uncanny valley stimuli can be seen to present mismatched features (Seyama and Nagayama, 2007; MacDorman et al., 2009; Mitchell et al., 2011; Saygin et al., 2012), such as a machine with a convincingly human voice, or an android with a physical appearance that is highly realistic but movements that are robotic. In this account, stimulus features, such as physical appearances, drive the automatic selection of a neural model for the purpose of predicting behavior. Stimulus mismatches can therefore lead to the selection of an inaccurate neural model, which is associated with error-related brain activity (Saygin et al., 2012), and error-related processing might result in negative affect. These neural models thus require learning in order to acquire ontological categories that subsequently produce contrasts due to feature mismatch.

THE UNCANNY VALLEY AS CATEGORICAL PERCEPTION

If feature mismatch is the result of the inclusion of features from neighboring categories, then a crucial feature of any general account of the UVH is the specification of category learning systems that acquire the category structure, as well as the representations that are retained within them (for a recent review, see Goldstone et al., 2012). A number of studies have attempted to qualify the UVH by making reference to principles and processes associated with categorization generally, and categorical perception more specifically (Cheetham et al., 2011, 2013; Moore, 2012; Burleigh et al., 2013; Yamada et al., 2013; Ferrey et al. submitted). Categorical perception (CP) accounts of the UVH suggest that this phenomenon is a consequence of categorical processes associated with stimulus identification. Specifically, stimuli along a human-likeness continuum are perceived as members of either a "human" or "non-human" category, except at the category boundary where their membership is ambiguous. This follows from the position that stimuli at the category boundary should not provide the observer with sufficient perceptual evidence to allow easy or accurate identification on the basis of their representation of the category structure. As a consequence, uncertainty and negative affect are produced due to competition during categorization response selection (Cheetham et al., 2011; Burleigh et al., 2013), which might in turn activate conflict resolution processes like inhibitory devaluation (Ferrey et al., submitted).

Empirical evidence is consistent with accounts of the uncanny valley based on categorical perception. For instance, Cheetham et al. (2011, 2013) demonstrated that participants' response latencies were longest when categorizing stimuli that were located at, or adjacent to, the category boundary on a human-avatar morph continuum. In addition to this, Burleigh et al. (2013, Experiment 2), Ferrey et al. (submitted), and Yamada et al. (2013), have each observed non-linear affective response functions across between-category (including human-animal and animal-animal) morph sequences that peaked at the midpoint between categories where stimuli were most ambiguous. Relative to the categorization literature, these accounts are underspecified, and therefore do not provide a complete account of the UCV phenomenon. Moreover, whereas Cheetham et al. (2011, 2013) and Yamada et al. (2013) have made a crucial connection between categorization and the response patterns associated with the uncanny valley, we cannot assume that categorization performance will be the only, or even the primary, determinant of affect. As we discuss, the uncanny valley might also be attributed to sub-categorical processes, such as those involved in assessing stimulus frequencies (Zajonc, 1968; Bornstein, 1989).

CATEGORY BOUNDARY AND EXEMPLAR REPRESENTATIONS

Any explanation of the UCV phenomenon based on categorical perception must consider categorization processes and representational assumptions (e.g., prototype-related models were recently considered by Moore, 2012). Most CP accounts of the UVH appear to have assumed that categorization is governed by a "category boundary" representation (Cheetham et al., 2011, 2013; Burleigh et al., 2013). Category boundary models suggest that when a stimulus is encountered, it is used to locate and modify the location of a decision boundary in perceptual space (Ashby and Gott, 1988). When individuals are presented with a novel stimulus, they will compare its location in perceptual space to

that of the category boundary. Proximity to the category boundary thereby increases categorization uncertainty (Paul et al., 2011; Schoenherr and Lacroix, 2014), and according to CP accounts of the UVH, proximity is also assumed to be inversely related to affect.

However, while a category boundary model might provide an adequate explanation of the uncanny "valley," which is a simple U-shaped quadratic function, it cannot account for the entire UCV response function, which is a more complex N-shaped cubic function (e.g., Mori, 1970). We suggest that models that take into consideration exemplar-based information might account for the additional features of a more complex function. Exemplar-based models assume that a memory trace is encoded each time a stimulus is encountered (Medin and Schaffer, 1978; Nosofsky, 1984). During the course of learning, each instance becomes associated with a category label, and at the end of learning each exemplar is represented by a probability distribution of features. Over the course of learning, an individual's attentional focus becomes reweighted to different regions of the stimulus continuum (Nosofsky, 1984, 1986), such that attention is sensitized to between-category differences and desensitized to within-category differences. When presented with a novel exemplar, individuals will compare it to all exemplars available in memory, and the similarity between the new item and old items in memory will determine the new item's category membership.

Thus, a key difference between category boundary and exemplar-based models is how individuals become sensitized to perceptual space. Category boundary models suggest that individuals can only typically become sensitized to a single region of perceptual space, namely where the category boundary is located; whereas exemplar-based models suggest that individuals can become sensitized to multiple regions of perceptual space, due to the distributions of individual members (Nosofsky, 1984, 1986).

THE UCV AS CATEGORICAL PERCEPTION OR FREQUENCY-BASED EXPOSURE

Crucially, affective processing of stimuli might not require the instantiation of categorical processes. The mere-exposure effect (Zajonc, 1968) suggests that repeated exposure to stimuli can lead to the formation of preferences, and negative affect might therefore be accounted for on the basis of familiarity or perceptual fluency alone (for a review, see Bornstein, 1989). In support of this, Harmon-Jones and Allen (2001) reported physiological evidence (via EMG and EEG) of affective responses that resulted from mere-exposure to stimuli, which corresponded with self-reported evaluations. If the mere-exposure effect can be extended to all members of a perceptual continuum, then an observer's familiarity with individual members of the continuum might be able to explain non-linear affective response functions. For example, along a human-likeness continuum that is anchored by "human" and "robot," individuals will have encountered a comparatively larger number of human instances relative to robots. Instances within these two categories should be much more familiar than instances that combine their features (e.g., androids). Thus, in contrast to the categorical perception account, a negative affective peak at an intermediate region in perceptual space might be explained by the fewer number of instances with the conjunction of features represented by stimuli in that region. On this basis, we suggest two distinct accounts of the UCV.

We suggest that at least two broad relationships are possible between cognitive and affective processing of stimuli, which we conceptualize as categorical perception (**Figure 2A**) and frequency-based exposure (**Figure 2B**) stage models. In conceptualizing these models, we limit ourselves to unidirectional processing. We assume that stimulus processing is mediated by the information that is stored in long-term memory, which includes memory traces of past episodes.

The categorical perception model (**Figure 2A**) reflects our understanding of extant categorical perception accounts of the UCV, in that it assumes categorical and affective responses derive from a common processing stage. In this model, individuals process sub-categorical information such as basic perceptual properties (e.g., stimulus magnitude, orientation) and frequency, but this information does not directly influence responding. Subsequent to this stage, category structures stored in long-term memory are activated, and these structures are used to determine both affective and categorical responses.

Alternatively, the frequency-based exposure model (**Figure 2B**) assumes that categorical and affective responses derive from separable processing stages. Specifically, affective responses are also driven by sub-categorical processing, which relies on frequency-based memory representations to provide more basic information such as frequency. The models defined in **Figures 2A,B** are sufficiently distinct that their predictions can be tested in a category-learning paradigm.

PRESENT STUDY

The present study was designed to test the predictions of the multistage models of the uncanny valley presented in **Figures 2A,B**, and a nested prediction concerning category structures. In the categorical perception model, the affective and categorical responses are derived from the same processing stage. Such a model therefore leads to a prediction of similar patterns of affective and categorical responses, as well as a strong and positive correlation between them. In contrast to this, the frequency-based exposure model implies that categorical and affective responses each account for unique sources of variance. Such a model therefore suggests that under some conditions patterns of responses might be similar, but they need not show a significant correlation.

Importantly, the stage models do not make predictions concerning the specific nature of categorical processing, only the relationship between categorical and affective responses. Therefore, a nested prediction concerns whether categorization will reflect category boundary or exemplar-based representations. The first possibility is that individuals will only have access to a category boundary representation that partitions the response continuum. Therefore, categorization accuracy and affective responses should increase, and response times should decrease, as a function of a stimulus' distance away from the category boundary. If participants are insensitive to individual characteristics, then categorization uncertainty should also be evidenced by a linear increase in response latencies as a function of proximity to the category boundary. Alternatively, if exemplar-based representations are

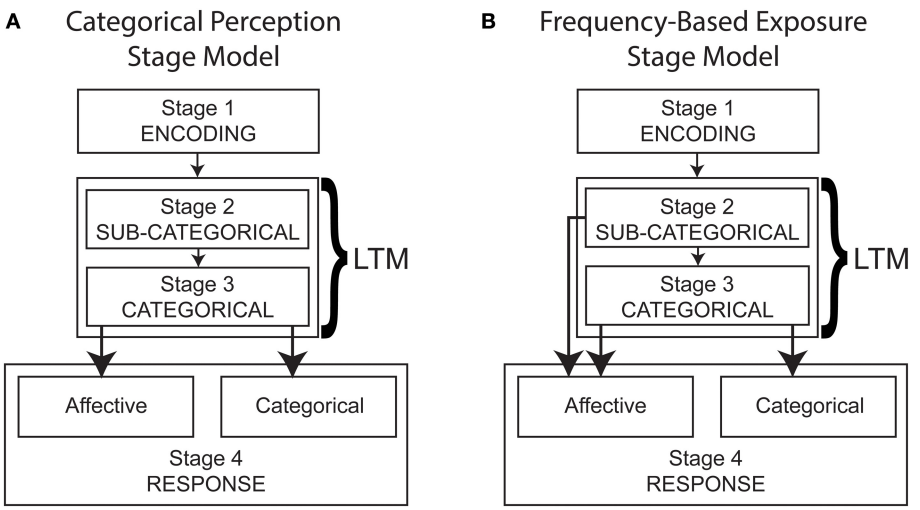


FIGURE 2 | Multistage UCV processing models derived from categorical perception (A) and frequency-based exposure accounts (B). Stage 1 (Encoding) relies on immediate memory whereas Stage 3 (Categorical) requires long-term memory (LTM) activation. Stage 2 (Sub-Categorical) requires activation of frequency-based properties retained in long-term memory.

Table 1 | Stimulus Frequencies for Training Session in Experimental Conditions for equal frequency, even distributions (EFED), unequal frequency, even distributions (UFED), and unequal frequency, uneven distributions (UFUD).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Tot.
EFED	–	–	4	4	4	4	4	–	4	4	4	4	4	–	–	40
UFED	–	–	8	6	4	2	–	–	–	2	4	6	8	–	–	40
UFUD	–	–	8	6	4	2	–	–	4	4	4	4	4	–	–	40

acquired for two contrasting categories and used for categorical processing, then the location of the central tendencies for each category should determine the location of the maxima and minima of the response functions for affective and categorization responses. Response latencies should evidence a similar trend. Specifically, if uncertainty in category membership is a function of exemplar frequency, then we would expect exemplars presented with comparatively high frequency during training to be associated with fast responses whereas exemplars presented with comparatively low frequency to be associated with slow responses.

In order to test the predictions of these models, our experimental design uses a category-learning paradigm in which we manipulate exemplar frequency along the perceptual continuum. Experiment 1 consists of two training conditions. In the first condition, stimuli within response categories are presented with equal frequency, with each category having an equivalent distribution (EFED). In the second condition, both category distributions are equivalent, but the exemplars were presented with unequal frequency (UFED) such that stimuli near the extrema of Categories A and B training sets are presented with the greatest frequency, and stimuli adjacent to the category boundary were presented with the lowest frequency. An important aspect of our design is that individuals are not exposed to the continuum extrema during the training phase. Thus, while the category boundary of the EFED and the UFED conditions should be identical, differences in exemplar frequency should decrease affective responses outside the training range if frequency-based information is a determinant of categorical and/or affective responses.

The results of Experiment 1 should provide a straightforward tests of our predictions. Left unaddressed, however, is what we consider to be a tacit property of UCV as discussed by Mori (1970): we are presented with less exemplar variability within one category (e.g., human) and greater exemplar variability in the contrasting category (e.g., non-human). In Experiment 2, we used one category defined by exemplars with equal frequencies selected from the EFED condition and another category defined by exemplars with unequal frequencies selected from the UFED condition. This procedure resulted in an unequal frequency, unequal distribution condition (UFUD) which we take as a closer approximation to the properties of the UCV first proposed by Mori (1970). **Table 1** provides training set frequencies.

Crucially, we were also interested in determining whether the affective response patterns could reasonably support a UCV interpretation. We distinguish between "strong" and "weak" interpretation as follows. The UCV function is a non-linear response function that is defined by a slope, indicating a category preference attributable to familiarity (e.g., for humans over robots), and a valley region that is located near the category boundary but skewed toward the preferred category. Thus, support for a strong interpretation of the UCV would be obtained if a response function possessed all of these features; support for a weak interpretation of the UCV would be obtained if a response function

possessed some of these features, such as a valley region without a slope. We anticipate the possibility that the EFED and UFED conditions might provide support for a weak interpretation, but not for the strong interpretation, due to their symmetry. In contrast, the UFUD condition might provide support for a strong interpretation of the UCV due to the asymmetry of the response function.

Although the stimuli that we use all represent non-human entities, we believe the findings of these studies are pertinent to human-like stimuli. By using non-human stimuli we hope to minimize the influence of stimulus familiarity or preference for human stimuli. This novelty facilitates the task of training participants to learn different category structures in an experimental setting with practical limitations (e.g., time). This manipulation also allows us to illustrate that response patterns associated with the UCV are generally patterns that can be attributed to stimulus familiarity rather than human-likeness, *per se*.

EXPERIMENT 1

Experiment 1 was designed as an initial test of our predictions derived from the hypothesized multistage models, and to provide evidence in support of the UCV phenomenon. We manipulated the frequency of stimulus presentation to differentially sensitize participants to regions of the stimulus response continuum. An equal frequency condition (EFED) was provided to half of the participants, wherein all stimuli within a category were presented with equal frequency, thereby creating a uniform distribution. An unequal frequency condition (UFED) was provided to the remaining half of the participants, wherein stimuli located within the middle of each category distribution were presented with higher frequency, thereby approximating a normal distribution. In each case, distributions of stimuli from Category A and Category B were symmetrical. Thus, by the end of training we hypothesized that participants should learn the distribution of the training stimuli equally well. Following training, participants responded to stimuli selected from the entire continuum. In the UFED condition, we additionally predicted that participants should show changes in affective responses due to less familiarity with the extreme values that in fact share fewer features with the contrasting category.

METHODS

Participants

A total of 60 participants were recruited online for this study (31 female, $M_{age} = 37.2$). Participants were recruited from Amazon's Mechanical Turk platform and paid a total of $5 if they completed all 4 sessions of the study ($1 for session 1, $1.25 for sessions 2 and 3, and $1.5 for session 4). All participants were registered with Mechanical Turk as United States residents. No participants reported having a visual impairment, and therefore no participants were excluded from our analyses. All participants consented to participate in the study.

Stimuli

Three morph sequences were generated, comprising the permutations of three distinct non-human creatures: a beast, a reptile, and an alien. These creatures were selected given our assumption that participants would have less familiarity with these categories thereby allowing us to more readily manipulate their frequency of exposure in the experimental context. Creatures were created using Daz Studio 4.6 Pro (daz3d.com) by modifying the morphology and texture of the *Genesis* base figure. Morph sequences were then created by stepwise adjustment of morphology and texture parameters corresponding to each creature. For example, the reptile creature had a "head scale" parameter which determined the size of its head, with a value of 32, whereas the alien creature had a value of 40. Therefore, the stimulus at the midpoint on the alien-reptile morph continuum assumed a value for this parameter that was half-way between the alien and reptile values (i.e., 36). Stimuli were then cropped in photo-editing software using an elliptical mask, and saved as images with a vertical resolution of 548 pixels. Stimuli were divided into training and test sets. The following stimuli were excluded from all training sets: stimulus 6 (the category boundary), and stimuli 1, 2, 14, and 15 (the extrema). Other stimuli were excluded depending on the frequency condition. For instance, stimuli 7 and 9 were not included in the training set for the UFED condition due to the frequency manipulation.

Procedure

Training. At the start of the experiment, participants were presented with stimuli during the training and test phases of the experiment by randomly assigning them to a creature continuum (for an example, see **Figure 3**) and a frequency condition (see **Table 1**). In order to control for the effect of creature continua, we used a counter-balanced design such that an equal number of participants were assigned to each of the (creature x frequency) conditions. This resulted in a total of 5 participants for each cell of the design, or 30 participants in each of the experimental conditions that were of interest. In the EFED condition, participants received an equal presentation of stimuli selected from the training range, whereas in the UFED condition participants received an unequal presentation of stimuli selected from the training range; in each case the frequency distributions were symmetrical.

At the beginning of training, participants were instructed that they would be presented with "models of unfamiliar living creatures" and that their task was to "learn what categories they belonged to." They were told that each creature was either a "Cax" or a "Miv" and that they were to press the "C" or "M" key depending on which type of creature they thought they saw. Participants were instructed to balance the demands of speed and accuracy. Key assignment was counter-balanced across participants.

Participants completed 1 training session per day over the course of 3 days. Each training session was composed of 10 blocks of 40 trials each, for a total of 400 trials per training session, and each session required approximately 20 min to complete. For each trial, a fixation point was presented for 500 ms, followed by a randomly selected stimulus from the training distribution for 750 ms (these timings were selected to be consistent with Cheetham et al., 2011). At the end of this sequence the response alternatives were presented until a response was registered. After a response was registered, feedback in the form of a "correct" or "incorrect" message was presented for 500 ms.

FIGURE 3 | Reptile-beast, beast-alien, and alien-reptile morph continua; stimuli shown here: 1, 3, 5, 7, 9, 11, 13, and 15.

Test. In the test session, all 15 stimuli were presented. Unlike the previous blocks, we sought to limit the amount of exposure to previously unseen stimuli. Therefore, the test session consisted of 4 blocks, in which each stimulus was presented 2 times each, for a total of 120 trials. The training session required approximately 12 min to complete. Stimulus presentation preceded in the same manner as in the training phase with two notable exceptions. Following presentation of a stimulus, participants were asked to rate its eeriness on a scale ranging from 1 (not at all eerie) to 7 (extremely eerie) using the "1" through "7" keys, respectively. After registering their response, participants were then asked to indicate whether it was a "Cax" or "Miv," as in previous sessions. The ordering of affective and categorization responses was deliberate in order to ensure that the effect of categorical information on ratings of eeriness would be limited.

Implementation. The study was developed for the web using HTML and JavaScript programming languages for the frontend, and PHP/MySQL for the backend. Preliminary tests using an automated responder on a test machine revealed that response time noise was within acceptable limits (i.e., less than 35 ms). Our online research environment is comparable to the one used by Crump et al. (2013). Crump et al., used JavaScript and recruited Mechanical Turk participants to successfully replicate numerous reaction time tasks like the Stroop (1935).

RESULTS

In order to test our predictions, we analyzed training and test responses separately in terms of categorization accuracy, response time, eeriness ratings, and the shape of categorization and affective response functions. A series of repeated-measures analyses of variance (ANOVA) were conducted. Greenhouse-Geisser adjusted values are reported with unadjusted degrees of freedom. All reported pairwise comparisons were conducted using a Bonferroni adjustment. We also report partial-eta squared as a measure of effect size. Following this, we use curve fitting analyses in order to facilitate our interpretation of the affective response functions.

TRAINING PHASE

Categorization accuracy

A repeated-measures ANOVA was conducted on categorization accuracy, using stimulus location relative to the category boundary (4) and response category (2) as within-subjects variables, and stimulus training distribution (2) as a between-subjects variable. Here, response categories (i.e., "Cax" and "Miv") were randomly assigned to stimuli located on the left- and right-halves of the stimulus continuum. Given the counterbalancing of stimulus sets, we collapsed across morph models prior to analysis. Stimuli directly adjacent to the category boundary in the UFED training condition were also removed prior to analysis. This adjustment was made due to the fact that these stimuli were not present in the EFED training condition and might introduce bias in the analysis. Similarly, the stimulus located at the category boundary was not presented during training and was therefore absent from the analysis of test responses.

Our analysis of training response accuracy revealed a significant main effect for stimulus distance from the category boundary, $F_{(3,\ 174)} = 11.921$, $MSE < 0.001$, $p < 0.001$, $\eta_p^2 = 0.17$. As **Figure 4** suggests, categorization accuracy increased as a function of distance from the category boundary with the lowest accuracy observed for stimuli nearest the category boundary (Stimuli 6 and 10) and the greatest accuracy observed for the most extreme stimuli (Stimuli 3 and 13). Supporting our interpretation of the data, pairwise comparisons revealed significant differences between stimuli nearest the category boundary (i.e., stimuli 6 and 10) and stimuli at all other distances ($ps < 0.012$). No other main effects or interactions reached significance, $ps > 0.1$[1]. Thus, the primary determinant of categorization accuracy during training was the location of a stimulus along the morphed continuum.

[1]A secondary analysis of arcsine transformed data revealed the same pattern as the analysis of untransformed data.

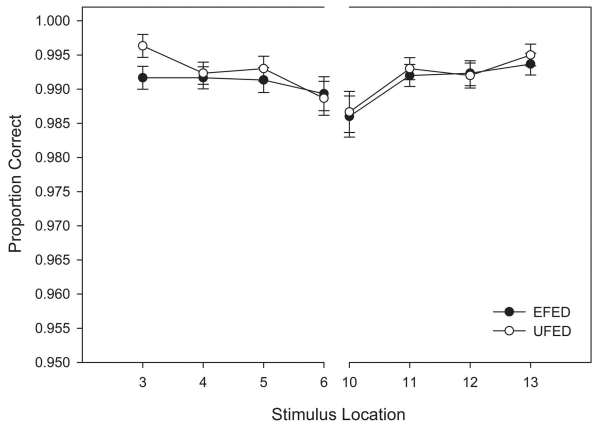


FIGURE 4 | Training categorization response accuracy in equal frequency (EFED) and unequal frequency conditions (UFED). Stimulus values correspond to stimuli selected from the training range (i.e., stimuli 3–13). Error bars represent 1 standard error of the mean ($N = 60$).

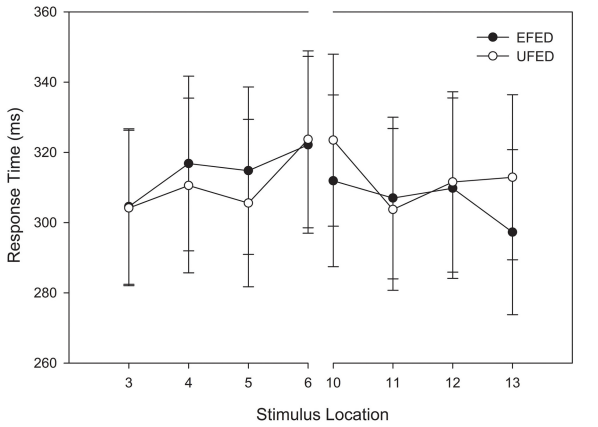


FIGURE 5 | Training categorization response time in equal frequency (EFED) and unequal frequency conditions (UFED). Stimulus values correspond to stimuli selected from the training range (i.e., stimuli 3–13). Error bars represent 1 standard error of the mean ($N = 60$).

Categorization response times

Using the same design as the analysis of accuracy, a repeated-measures ANOVA was conducted on categorization response time. To eliminate outlying observations, we first computed an unadjusted mean response time for each participant and identified trials wherein their responses were 3 standard deviations above the mean. This accounted for 2.1% of trials. Consistent with our analysis of categorization accuracy, categorization response time decreased as a function of distance from the category boundary, $F_{(3,\ 174)} = 5.061$, $MSE = 1380.276$, $p = 0.005$, $\eta_p^2 = 0.08$.

As **Figure 5** demonstrates, stimuli near the category boundary were associated with longer response times than stimuli at more distal locations. Pairwise comparisons revealed that the response latency for stimuli close to the category boundary significantly differed for adjacent stimuli and those located at extreme distances (ps < 0.044). No other main effects or interactions reached significance, ps > 0.1. Again, these results provide additional evidence that the primary determinant of categorization response time during training was the location of a stimulus along the morphed continuum.

TEST PHASE

Response accuracy

A repeated-measures ANOVA was conducted on accuracy of responses obtained during the test phase. This analysis was comparable to that of the training phase with the exception that due to the uniform distribution used during the test phase for both training groups, stimuli adjacent to the category boundary as well as novel extrapolation items outside the range of the training items were also included in the analysis. Again, the stimulus located at the category boundary was eliminated from the analysis due to its ambiguity (it was entered into another analysis, see the Categorization response times and Affective ratings of eeriness sections below). Thus, stimulus location relative to the category boundary (7) and response category (2) were entered as within-subjects variables, and stimulus training distribution (2) was entered as a between-subjects variable.

Replicating the findings of categorization accuracy obtained in the training phase, we observed a significant main effect of stimulus distance from the category boundary, $F_{(6,\ 348)} = 54.516$, $MSE = 0.021$, $p < 0.001$, $\eta_p^2 = 0.485$. An interaction was also observed between stimulus distance and frequency distribution, $F_{(6,\ 348)} = 5.292$, $MSE = 0.021$, $p = 0.007$, $\eta_p^2 = 0.08$[2]. An examination of **Figure 6** reveals a more pronounced decrement in categorization accuracy around the category boundary in the test phase relative to the training phase. This trend was especially pronounced for participants in the unequal frequency (UFED) training condition, and suggests that participants were affected by the distributional properties in the test phase.

Categorization response times

A repeated-measures ANOVA was conducted on categorization response time in the test phase. In order to compare to affective response functions (see below), the stimulus at the category boundary was also included. Therefore, unlike previous analyses, the entire stimulus continuum was tested. Thus, stimulus location (15) was entered as a within-subjects variable and stimulus training distribution (2) as a between-subjects variable. An analysis of response time outliers was again conducted on individual participants' responses. After obtaining an unadjusted mean, no responses were observed to be larger than 3 standard deviations above the mean. This result is not surprising given the reduced number of replications in the test phase.

As with the response time analysis in the training phase, a main effect was observed for stimulus location, $F_{(14,\ 812)} = 4.631$, $MSE = 188391.325$, $p = 0.002$, $\eta_p^2 = 0.053$. **Figure 7** indicates that this effect can be accounted for by the slower response times for stimuli that were at, and adjacent to, the category boundary.

[2]A secondary analysis of arcsine transformed data revealed the same pattern as the analysis of untransformed data.

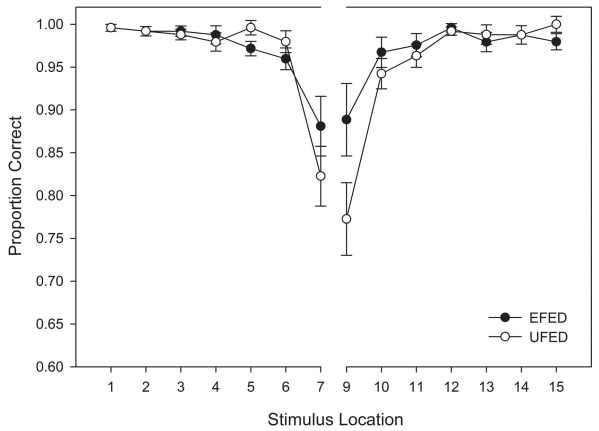


FIGURE 6 | Test categorization response accuracy in equal frequency (EFED) and unequal frequency conditions (UFED). Stimulus values correspond to stimuli selected from the entire stimulus range (i.e., stimuli 1–15). Error bars represent 1 standard error of the mean ($N = 60$).

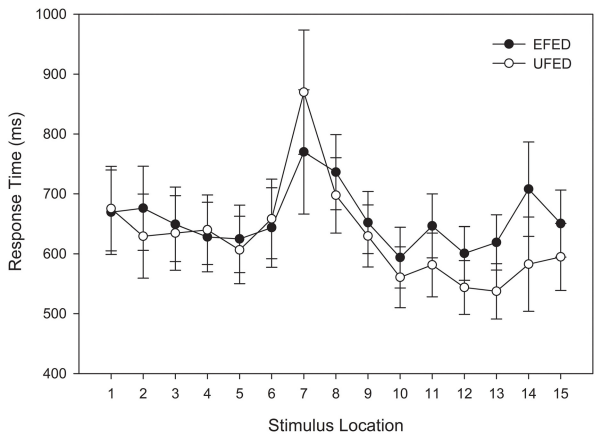


FIGURE 7 | Test categorization response time in equal frequency (EFED) and unequal frequency conditions (UFED). Stimulus values correspond to stimuli selected from the entire stimulus range (i.e., stimuli 1–15). Error bars represent 1 standard error of the mean ($N = 60$).

Category response frequencies

In the categorical perception literature, category boundaries were originally assessed by examining category response frequencies across a stimulus continuum (cf. Pisoni and Tash, 1974). These analyses allow for the identification of the location and shape of a category boundary. Whereas continuous increases in stimulus magnitude relative to some criterion (e.g., brightness, size) can be fit a continuous function, categorical perception is typically reflected in a sigmoidal function (Harnad, 1987). Our analyses determined the frequency of "Cax" responses as a proportion of total category responses for each level along the stimulus continuum (e.g., frequency of "Miv" responses reflects the inverse of this function). These results were then plotted across the stimulus continuum, and a sigmoid function was fitted to the data.

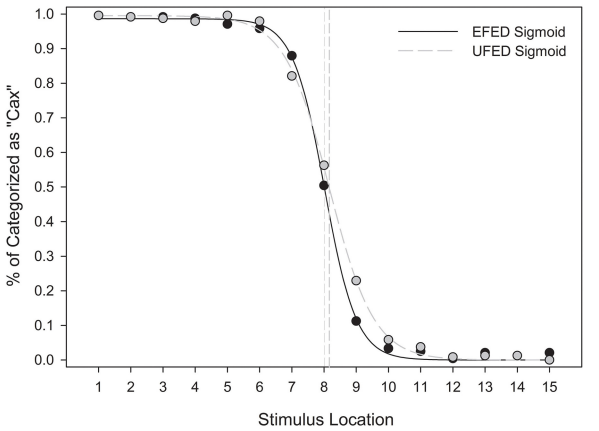


FIGURE 8 | Mean categorization response frequencies for equal frequency (EFED) and unequal frequency conditions (UFED). Stimulus values correspond to stimuli selected from the entire stimulus range (i.e., stimuli 1–15). Vertical dashed lines represent the estimated category boundaries.

Figure 8 suggests that the category response frequencies in EFED and UFED conditions were consistent with a sigmoidal shape, and indicate that a category boundary was present at or near stimulus 8, the mid-point of the stimulus continuum. The sigmoid function provided an adequate fit in the EFED $[F_{(2,\ 12)} = 7943.437, MSE < 0.001, p < 0.001, R^2_{adj} > 0.999]$ and UFED $[F_{(2,\ 12)} = 8110.364,\ MSE < 0.001,\ p < 0.001,\ R^2_{adj} > 0.999]$ conditions. Parameter estimates confirm that the point of inflection in each case was approximately located at stimulus 8 ($x_{0,\mathrm{EFED}} = 8.020$, $x_{0,\mathrm{UFED}} = 8.165$). Thus, stimulus identification in the test phase is consistent with categorical perception.

Affective ratings of eeriness

In order to examine the location and property of the global minima in eeriness ratings that correspond to the uncanny valley, we conducted a repeated-measures ANOVA that included stimulus location (15) as a within-subjects variable and stimulus training distribution (2) as a between-subjects variable. Relative to the previous analyses of categorization accuracy and response time, this approach allows for a straightforward comparison between the shapes of the function that fit affective responses provided below. Our analysis revealed a marginally significant effect of stimulus location, $F_{(14,\ 812)} = 3.267$, $MSE = 15.128$, $p = 0.055$, $\eta^2_p = 0.053$. The interaction between stimulus location and training distribution did not approach significance, $F_{(14,\ 812)} = 1.706$, $MSE = 15.128$, $p = 0.193$, $\eta^2_p = 0.029$.

Although an interaction was not observed between stimulus location and stimulus training distribution, **Figure 9** suggests that the trend of affective responses did change as a function of stimulus training distribution. In the EFED condition, an overall linear trend was observed across the continuum, with an affective minimum at one end of the stimulus continuum, and an affective maximum at the other end. Such a pattern would be expected if participants were using one response category as reference point and comparing stimulus exemplars to that category. By contrast,

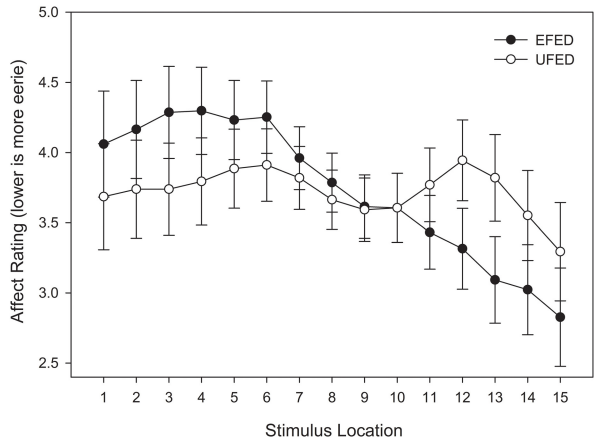


FIGURE 9 | Test affective responses (lower is more eerie) in equal frequency (EFED) and unequal frequency conditions (UFED). Stimulus values correspond to stimuli selected from the entire stimulus range (i.e., stimuli 1–15). Error bars represent 1 standard error of the mean ($N = 60$).

in the UFED condition an M-shape was instead observed, with affective minima at the category boundary as well as at the endpoints of the stimulus continuum. When comparing these results to those obtained in the analyses of categorization responses, it is instructive to note that the stimuli associated with the slowest response times and lowest levels of categorization accuracy were not those that generated the highest levels of eeriness. As such, it is reasonable to conclude that there is a degree of dissociation between categorization performance and affective responses. Thus, while categorical perception appears to be compatible with the uncanny valley hypothesis, it also appears that the affective component of the valley is influenced by other affective and cognitive processes.

Correlations

The variation in patterns across our dependent measures prompted an examination of the relationship between these measures. Test accuracy, test response time, and affective ratings were included in a correlational analysis. We examined the EFED and UFED conditions separately.

As **Table 2** indicates, in both the EFED and UFED conditions, marginally significant negative correlations were obtained between response time and categorization accuracy, $r_{(14)} = -0.485$, $p = 0.079$, and $r_{(14)} = -0.517$, $p = 0.059$, respectively. However, in both conditions, the correlation between categorization response time and eeriness did not reach significance, $ps > 0.5$. Thus, while an increase in response time was observed near the category boundary, the remaining differences in responses did not support an interpretation that response time and eeriness ratings were produced by the same response processes. Equally important, in both conditions, the correlation between accuracy and eeriness did not reach significance, $ps > 0.5$. Thus, while it appears that information processing associated with the production of a categorization response and affect were related, category membership and affective responses differed in important ways. These differences appear to be a result of novel extrapolation items, something that is inconsistent with a category boundary model of the UCV.

Table 2 | Pearson correlations of dependent measure in the test phase for equal frequency (EFED) and unequal frequency conditions (UFED).

		Response Time	Accuracy
EFED			
	Accuracy	−0.485 (0.079)	-
	Eeriness	0.076 (0.796)	−0.059 (0.842)
UFED			
	Accuracy	−0.517 (0.059)	-
	Eeriness	0.136 (0.643)	0.076 (0.797)

p-values are in brackets.

Curve fitting analysis

Mori's original proposal assumed that the uncanny valley is characterized by a non-linear response function. In the present experiment, we sought to directly test this assumption by fitting curves to the obtained response functions (see also Burleigh et al., 2013) for both EFED and UFED training conditions. A second goal of the present analysis was to obtain evidence for the underlying representation that supports the uncanny valley either in terms of a category boundary or an exemplar-based representation.

A number of non-linear functions were selected on theoretical grounds that were not included in Mori's original characterization of the model. In particular, our manipulation of frequency effects in the context of a categorization experiment was motivated by the belief that when participants are sensitized to specific regions of the response continuum, the location of affective minima and maxima can be manipulated. As we noted, a category boundary representation would be evidenced by a U-shaped quadratic function, whereas an exemplar-based category representation would be evidenced by an M-shaped quintic function.

We used a curve fitting analytic approach to test these possibilities, by fitting polynomials of degree 0 through 5 (i.e., constant, linear, quadratic, cubic, quartic, and quintic) to the means. Curve fitting was performed using Origin Lab (originlab.com) software. We used the Akaike Information Criterion (AIC; see Burnham and Anderson, 2002) as our goodness-of-fit index. The AIC is suited to comparing models with different degrees of complexity because it penalizes models with additional fit parameters. We calculated raw Akaike values and Akaike Weights (w_i), which are a transformation of raw scores that indicate the probability that a particular model among the set of models is correct (Wagenmakers and Farrell, 2004). Using these weights, we also calculated evidence ratios by dividing the weight of one model by the sum of all weights. These ratios are understood in context of a "confidence set," which is similar to a confidence interval and is defined as 10% of the highest Akaike Weight in the set (Royall, 1997). Thus, models falling outside of the confidence set can be rejected as poorer fits to the data. For the purposes of interpretation, it should be noted that lower raw Akaike values and higher Akaike Weights indicate a better fit

to the data. The results of these analyses are summarized in **Table 3**.

In the EFED condition, the constant, linear, and quadratic models fell outside the confidence set. Thus, accounts based on random responding or based on the association of equivalent negative affect for each stimuli are not supported (constant). Participants also did not appear to be solely biased by one end of the response continuum (linear). Perhaps most importantly for our purposes, a failure to obtain a fit for the quadratic function suggests that a category boundary model does not provide a good fit to the data in the absence of other assumptions. Instead, the model within the confidence set that was most likely to represent the data was the quartic model, $w_i(AIC) = 0.542$. Thus, there is a 54.2% chance that the quartic model best accounts for the pattern or data we observed. However, the cubic model obtained an Akaike weight of a similar magnitude, $w_i(AIC) = 0.396$. Given the similarity of these model weights, it would be reasonable to select the cubic function over the quartic function on the basis of its parsimony. This finding suggests that, for at least one category, the stimulus frequency manipulation produced a change in response affect and that a category boundary model cannot adequately account for the data.

In the UFED condition, we again observed that the constant and linear models fell outside the confidence set. In the same manner as the EFED condition, this suggests that a constant response bias or uniform negative affect were not evidenced in participants' responses (constant) and that a single category was not used as the sole basis for comparison (linear). Instead, the model within the confidence set that is most likely to represent the data was the quintic model. $w_i(AIC) = 0.693$, and the next best model was the quartic model, $w_i(AIC) = 0.131$. Thus, the obtained difference between these models clearly suggests that a quintic function best represents this data set. Taken along with the EFED results, the observation that a quintic function provides the best fit again suggests that the inclusion of exemplar-based representation is an important feature of a model of the UCV phenomenon.

Table 3 | Residual sums of squares (RSS) and Akaike values for equal frequency, equal distribution (EFED), unequal frequency conditions, equal distribution (UFED), and unequal frequency, unequal distribution (UFUD) conditions.

Training Set	Model	*RSS*	*AICc*	$\Delta_i(AIC)$	$w_i(AIC)$	*CI*
EFED Condition	Constant[1]	3.521	−19.43	52.81	< 0.001	0.054
	Linear[2]	0.431	−48.26	23.98	< 0.001	–
	Quadratic[3]	0.152	−60.70	11.54	0.002	–
	Cubic[4]	0.057	−71.61	0.63	0.396	–
	Quartic[5]	0.040	−72.24	0.00	0.542	–
	Quintic[6]	0.036	−67.84	4.40	0.060	–
UFED Condition	Constant[1]	0.386	−52.59	4.76	0.064	0.069
	Linear[2]	0.328	−52.36	4.99	0.057	–
	Quadratic[3]	0.246	−53.45	3.90	0.099	–
	Cubic[4]	0.235	−50.36	6.99	0.021	–
	Quartic[5]	0.135	−54.01	3.34	0.131	–
	Quintic[6]	0.073	−57.35	0.00	0.693	–
UFUD Condition	Constant[1]	1.145	−36.28	23.26	< 0.001	0.069
	Linear[2]	0.534	−45.03	14.51	< 0.001	–
	Quadratic[3]	0.188	−57.52	2.02	0.252	–
	Cubic[4]	0.181	−54.22	5.32	0.049	–
	Quartic[5]	0.093	−59.54	0.00	0.693	–
	Quintic[6]	0.091	−53.99	5.54	0.043	–

Superscript denotes K, the number of parameters in the model.

DISCUSSION

The results of Experiment 1 add further evidence to the literature for the existence of the UCV phenomenon. Our results, however, qualify categorical perception accounts. Our measures of categorization accuracy, response frequencies, and response latencies all produced categorical response functions indicating that participants successfully learned the category structures. These analyses also suggest that the equivalent categorization performance was obtained on either side of the category boundary, which indicates that categories (i.e., Cax and Miv) were learned equally well. Similarly, whether participants were trained with exemplars with equal frequencies or unequal frequencies did not appear to alter the location of the category boundary. In both equal frequency (EFED) and unequal conditions (UFED), the category boundary was located at stimulus 8. Greater accuracy was obtained for items adjacent to the category boundary in the equal frequency condition relative to those in the unequal frequency condition. While such findings indicate categorical perception, it is not necessarily the case that categorical processing is the primary determinant of affective responses.

In order for the UCV to be understood in a manner similar to Mori's initial conceptualization, an affective relationship must be established with the location of exemplars along a continuum. Our curve-fitting analysis of eeriness ratings indicated that there were differences between the frequency training conditions. In the equal frequency condition, the response pattern was best fit by a cubic function. This pattern was evidenced by a slope, indicating that one response category was preferred to the other, and also a non-linear component at one end of the stimulus continuum. This pattern did not conform to our *a priori* hypotheses. Therefore, we can only speculate about its causes. One possibility is that the response pattern was an artifact of the category-response-key mappings that were used in our design. Each response category was assigned to a specific key on the keyboard, and participants were instructed to use index fingers on different hands for each key. As the location of the "C" and "M" keys are fixed on a standard QWERTY keyboard, handedness could have played a role. A second possibility is that the response labels themselves could have introduced some bias. For instance, participants might have preferred "Cax" because it occurs earlier in the alphabet, or because it was more familiar to them (due to associations with phonetically similar words), or they might have adopted a related response heuristic where in one category was used to anchor judgments (e.g., due to reading labels from left to right), or one of the category labels might have been more meaningful than another which resulted

in differential leaning outcomes (as has been observed with nonsense syllables, see Davis, 1930). If these factors systematically affected performance, they do not appear to be evidenced in the unequal frequency condition. The asymmetries obtained in the EFED condition are likely a result of idiosyncratic response biases and strategies used by the participants in this condition. In the unequal frequency condition, the response pattern was best fit by a quintic function. This pattern was M-shaped, with a valley-region located near the category boundary, and two affective minima located near the extrema. Importantly, however, because the pattern was symmetrically distributed around the category boundary, it did not possess all of the features of the classic UCV function and therefore would only support a weak UCV interpretation.

The differences in response functions for categorization accuracy and eeriness ratings also suggest an important relationship that has been neither specified nor explicitly examined in the literature examining the UVH. Namely, we found that categorization accuracy and affect did not significantly co-vary. Such a finding has important implications for studies of the UVH that claim that it can be accounted for by categorical perception. We suggest that the methods of Experiment 1 played a key role in dissociating affective and categorical responses. By requiring affective responses immediately after stimulus presentation and prior to categorical responses, the probability that categorical information was available was reduced. Two differences in findings provide clear demonstrations of this dissociation. First, the small amount of error variance in the categorization responses observed for the items on the end of the distribution in the test session can be sharply contrasted against the larger error variance for the affective responses. Second, whereas categorization accuracy was uniformly high and response times were uniformly fast for items located near the ends of the distributions, affective ratings instead showed asymmetric effects with the response functions.

Another interesting finding was the absence of a relationship between response time and eeriness. Long response latencies are typically taken as evidence of response uncertainty. If eeriness is a consequence of uncertainty in the category membership of an exemplar, then eeriness and response times should exhibit a positive correlation. Instead, the absence of a significant correlation suggests that the processes that determine the uncertainty in category membership and the processes underlying eeriness might be supported by different affective and cognitive processes. Thus, whereas the uncanny valley appears to be a product of experience with exemplars, these two processes appear to be separable. It is necessarily the case that at some level of processing these processes must be influenced by the same stimulus information. Yet stages of processing appear to be evidenced such that affective ratings were influenced more by novel exemplars that represented extrapolations for the range of training stimuli, whereas categorization responses appear to be primary influenced by categorical representations of the stimulus continuum stored in long-term memory.

Experiment 1 therefore provides preliminary evidence in support of the frequency-based exposure model of the UVH, and against the categorical perception model. However, categorization responses were consistent with a category boundary representation. Therefore, we suggest that frequency-based memory representations and category boundary representations are both stored in long-term memory, but that the representation of this information produces different patterns of performance in affective and categorical responses. A remaining possibility is that the UCV phenomenon requires both unequal frequencies within a category, and unequal distributions for both reference categories. Experiment 2 examines this possibility.

EXPERIMENT 2

Experiment 2 was conducted to clarify the relationship between variables observed in Experiment 1 while also further investigating the effect of distributional properties on categorical and affective responses. One way to interpret Mori's (1970) proposal is that there is nothing intrinsically important about the human category. Rather, it is only our frequency of exposure to, or familiarity with, stimuli that results in a category being used as a point of reference. As a result, non-human categories contrasted against the human category are likely to be perceived as less familiar due to their lower frequency. Thus, Mori's proposal might take for granted that two conditions need to be met for the experience of eeriness to occur when contrasting categories: a small number of items from one category need to be observed with high frequency and a larger number of items need to be observed from a contrasting category with unequal frequency.

METHODS

Participants

A total of 30 participants were recruited online for this study (12 female, $M_{age} = 34.3$). As before, participants were recruited from Amazon's Mechanical Turk platform and paid a total of $5 if they completed all 4 sessions of the study. All participants were registered with Mechanical Turk as United States residents. No participants reported having a visual impairment, and therefore no participants were excluded from our analyses. All participants consented to participate in the study.

Stimuli

The same stimuli were used to ensure a direct comparison with Experiment 1.

Procedure

All procedures were identical to Experiment 1 with the exception of training frequency. In Experiment 2, we presented participants with one category from the equal frequency condition in Experiment 1 and another category from the unequal frequency condition. We refer to this distribution as the unequal frequency, unequal distribution (UFED) condition below.

RESULTS

As in Experiment 1, we analyzed categorization accuracy, categorization response times, categorization response frequencies, and affective responses in the training and test phases. We also conducted curve fitting analyses to facilitate our interpretation of the affective response pattern. Again we report Greenhouse-Geisser unadjusted values and unadjusted degrees of freedom.

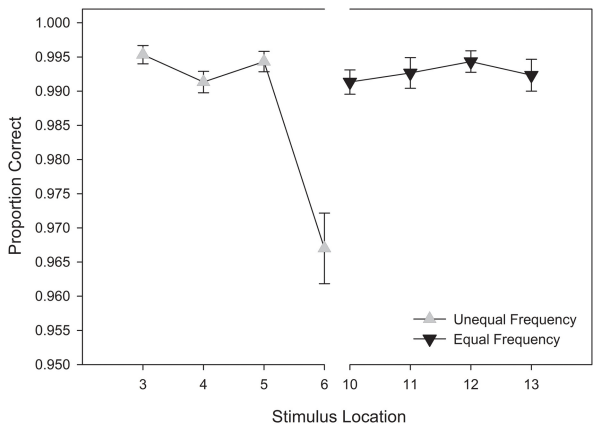


FIGURE 10 | Training categorization response accuracy in the unequal frequency, unequal distribution condition. Stimulus values correspond to stimuli selected from the training range (i.e., stimuli 3–13). Error bars represent 1 standard error of the mean ($N = 60$).

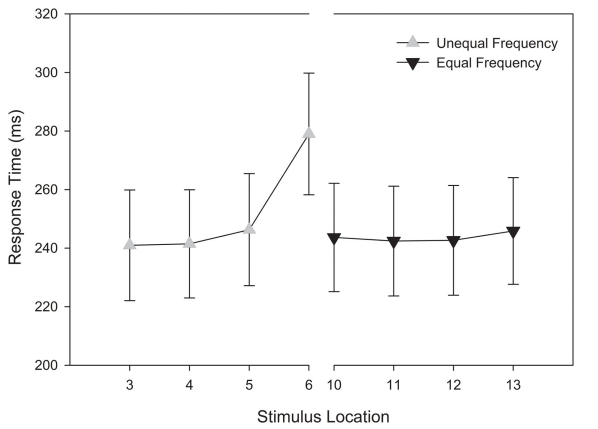


FIGURE 11 | Training categorization response time in the unequal frequency, unequal distribution condition. Stimulus values correspond to stimuli selected from the training range (i.e., stimuli 3–13). Error bars represent 1 standard error of the mean ($N = 60$).

TRAINING PHASE

Response accuracy

A repeated measures ANOVA was performed on categorization accuracy, using stimulus location relative to the category boundary (4) and response category (2) as within-subjects variables. Unequal and equal frequency categories were assigned to stimuli located on the left- and right-halves of the stimulus continuum, respectively. Similarly, we also collapsed across morph models and excluded the category boundary stimulus, as well as stimuli directly adjacent to the category boundary, because training for these stimuli occurred for only one of the two response categories. As in Experiment 1, we obtained a significant main effect for stimulus distance from the category boundary, $F_{(3,\ 87)} = 21.725$, $MSE < 0.001$, $p < 0.001$, $\eta_p^2 = 0.428$. However, unlike Experiment 1, we also obtained a significant main effect for category, $F_{(1,\ 29)} = 8.942$, $MSE < 0.001$, $p = 0.006$, $\eta_p^2 = 0.236$, and a significant interaction between category and stimulus distance, $F_{(3,\ 87)} = 20.951$, $MSE < 0.001$, $p < 0.001$, $\eta_p^2 = 0.419$[3].

Figure 10 suggests that this interaction can be accounted for by an asymmetry in the frequency of exemplars contained within the response categories. Specifically, with responses to the unequal frequency category, the stimulus nearest the category boundary (stimulus 6) was associated with lower categorization accuracy than stimuli at more distal locations. In contrast, the stimulus nearest the boundary (stimulus 10) in the equal frequency category received accuracy similar to other stimuli in its response category. These results replicate the unequal and equal frequency conditions in Experiment 1, respectively.

Response times

A similar repeated measures ANOVA was conducted for categorization response time. Prior to this analysis, we removed outlying trials with response times greater than 3 standard deviations from a participant's mean response time for each stimulus. This resulted in a removal of 2% of all trials. As with the analysis of accuracy, main effects were observed for exemplar distance from the category boundary, $F_{(3,\ 87)} = 9.048$, $MSE = 780.019$, $p < 0.001$, and response category, $F_{(1,\ 29)} = 6.481$, $MSE = 635.357$, $p = 0.016$, as well as an interaction between stimulus distance and response category, $F_{(3,\ 87)} = 13.83$, $MSE = 538.695$, $p < 0.001$.

Figure 11 suggests that, as in the case of accuracy, this interaction can be accounted for by an asymmetry in the response categories. Specifically, in the unequal frequency condition, the stimulus nearest the category boundary (stimulus 6) was associated with slower response times than were stimuli at more distal locations. In contrast, stimulus 10 in the equal frequency condition was associated with response times equivalent to those of other stimuli in its response category.

TEST PHASE

Response accuracy

A repeated measures ANOVA was also conducted on categorization response accuracy for stimuli presented during the test phase. As in Experiment 1, this analysis included stimuli adjacent to the category boundary, because participants received an equal frequency of these stimuli in each response category during test. Again, the category boundary stimulus was not included, as there was no objective criteria that could be used to determine accuracy. Thus, stimulus location relative to the category boundary (7) and response category (2) were entered as within-subjects variables.

Replicating the findings of the training phase, we observed significant main effects for stimulus distance from the category boundary, $F_{(6,174)} = 39.887$, $MSE = 0.027$, $p < 0.001$, $\eta_p^2 = 0.579$, response category, $F_{(1,\ 29)} = 4.516$, $MSE = 0.021$, $p = 0.042$, $\eta_p^2 = 0.135$, and a significant interaction between response category and stimulus distance, $F_{(6,\ 174)} = 4.082$, $MSE = 0.051$,

[3]A secondary analysis of arcsine transformed data revealed the same pattern as the analysis of untransformed data.

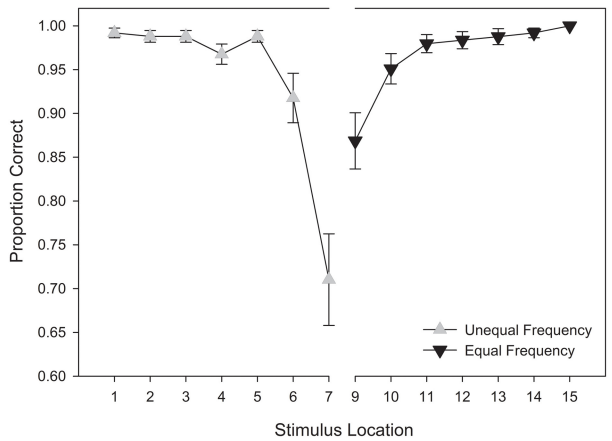


FIGURE 12 | Test categorization response accuracy in the unequal frequency, unequal distribution condition. Stimulus values correspond to stimuli selected from the training range (i.e., stimuli 3–13). Error bars represent 1 standard error of the mean ($N = 60$).

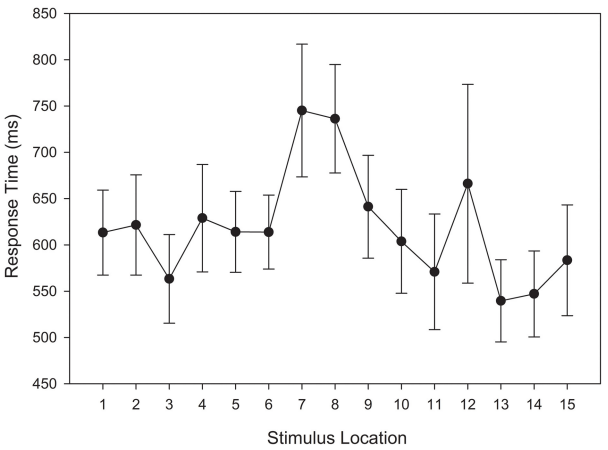


FIGURE 13 | Test categorization response time in the unequal frequency, unequal distribution condition. Stimulus values correspond to stimuli selected from the entire stimulus range (i.e., stimuli 1–15). Error bars represent 1 standard error of the mean ($N = 60$).

$p = 0.036$, $\eta_p^2 = 0.134$[4]. An examination of **Figure 12** indicates that in each response category, the boundary-adjacent stimulus was associated with lower accuracy. However, accuracy was lower for the boundary-adjacent stimulus in the unequal frequency category in comparison to the equal frequency category.

Categorization response times

A repeated measures ANOVA was conducted with categorization response time as the dependent variable. Unlike the analysis of training stimuli, no outliers met the removal criterion. Therefore, all responses were entered into the response time analysis. A main effect for stimulus distance from the category boundary reached significance, $F_{(6,\ 174)} = 3.155$, $MSE = 40747.27$, $p = 0.044$, $\eta_p^2 = 0.098$. **Figure 13** indicates that response times for stimuli adjacent to the category boundary were relatively slower than response times for other stimuli. Again, this reflects uncertainty in category membership.

Category response frequencies

As in Experiment 1, we determined the frequency of "Cax" responses as a proportion of total category responses for each level along the stimulus continuum. These response frequencies were plotted across the stimulus continuum, and fitted a sigmoid function to the data.

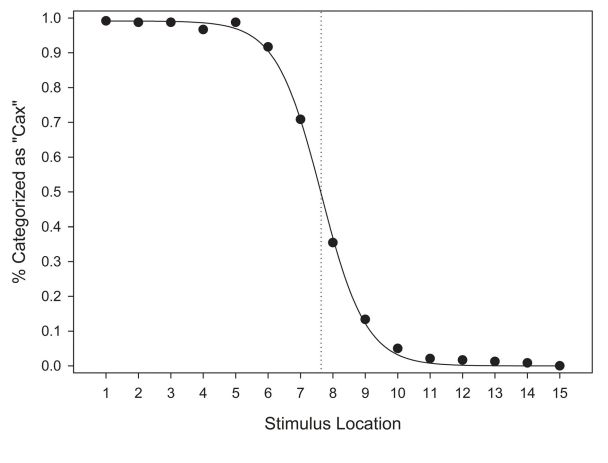


FIGURE 14 | Categorization response frequencies in the unequal frequency, unequal distribution condition. Stimulus values correspond to stimuli selected from the entire stimulus range (i.e., stimuli 1–15). The vertical dashed line represent the estimated category boundary.

Figure 14 suggests that the category response frequencies were consistent with a sigmoidal function, and indicate that a category boundary was present near stimulus 8. A sigmoid function was found to provide an adequate fit to the data, $F_{(2,\ 12)} = 8312.653$, $MSE < 0.001$, $p < 0.001$, $R^2_{adj} > 0.999$. Parameter estimates indicate that the point of inflection was located at 7.63. These results suggest that, in comparison to the EFED and UFED conditions in Experiment 1, the location of the category boundary was biased toward the unequal frequency category.

In order to test the possibility that the category boundary in the UFUD condition was biased due to the unequal frequency category, we conducted a follow-up analysis. We fit a sigmoid function to the response frequencies for each participant in all three of the training conditions. Then, we used two-tailed *t*-tests

[4]A secondary analysis of this data was conducted using arcsine transformed proportion correct values. This analysis revealed a similar pattern as the analysis of untransformed data. A change from significant to marginally significant results were observed for the main effect of response category [$F_{(1,\ 29)} = 4.161$, $MSE = 0.058$, $p = 0.051$, $\eta_p^2 = 0.125$] and also for the interaction between response category and stimulus distance [$F_{(6,\ 174)} = 2.666$, $MSE = 0.093$, $p = 0.074$, $\eta_p^2 = 0.084$]. The differences between these analyses of transformed and untransformed data can likely be attributed to a reduction in observed power. Specifically, power for the main effect of response category was reduced from $\beta = 0.538$ to $\beta = 0.505$, and power for the interaction was reduced from $\beta = 0.602$ to $\beta = 0.527$.

to compare the mean inflection points for each of the training conditions. As expected, the location of the category boundary did not differ between the EFED ($M = 7.955$, $SD = 0.586$) and UFED ($M = 8.186$, $SD = 0.786$) conditions, $t_{(58)} = 1.283$, $p = 0.205$. However, a significant difference was observed between the UFUD ($M = 7.641$, $SD = 0.815$) and UFED conditions, $t_{(58)} = 2.636$, $p = 0.011$; and a marginally significant difference was observed between the UFUD and EFED conditions, $t_{(58)} = 1.713$, $p = 0.092$.

Affective ratings of eeriness

A repeated measures ANOVA was conducted on affective ratings, with stimulus location (15) entered as a within-subjects variable. The main effect of stimulus location did not reach significance, $F_{(14,\ 406)} = 1.387$, $MSE = 15.306$, $p = 0.258$, $\eta_p^2 = 0.046$. However, a visual inspection of **Figure 15** suggests a pattern that is aligned with our expectations. Specifically, there appears to be a local minimum near the category boundary, and a slope that indicates a small bias toward the unequal frequency category.

Correlations

Test accuracy, test response time, and affective ratings were included in a correlational analysis. Replicating the results of Experiment 1, we again observed a significant correlation between response time and categorization accuracy, $r_{(14)} = -0.78$, $p < 0.001$. Neither the correlation of eeriness ratings and accuracy, $r_{(14)} = -0.14$, $p = 0.64$, nor eeriness ratings and categorization response time reached significance, $r_{(14)} = 0.26$, $p = 0.36$. As in Experiment 1, these findings can be taken as suggesting that the response processes associated with affective ratings and categorization differ.

Curve fitting analysis

Next, we were interested in assessing the affective response trend. A visual inspection of **Figure 15** suggests that the UFUD condition produced a distorted M-shape with a shallower region of negative affect near the category boundary, and a positive bias toward the unequal frequency category. As in Experiment 1, we fit polynomials of degree 0 through 5 (i.e., constant, linear, quadratic, cubic, quartic, and quintic) to the data, and we used the AIC as our goodness-of-fit index when comparing models.

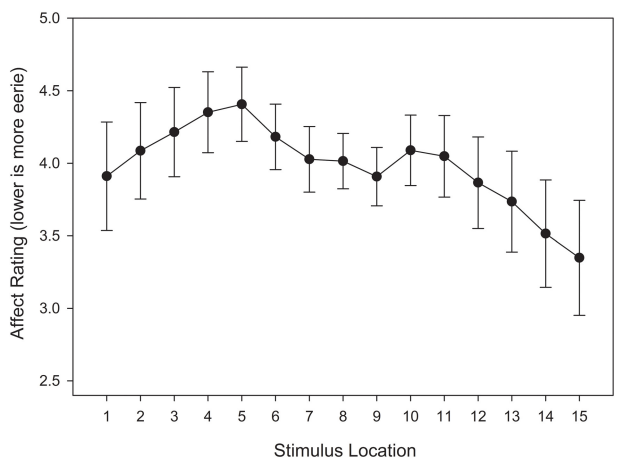


FIGURE 15 | Test affective responses (lower is more eerie) in the unequal frequency, unequal distribution condition. Stimulus values correspond to stimuli selected from the entire stimulus range (i.e., stimuli 1–15). Error bars represent 1 standard error of the mean ($N = 60$).

As **Table 3** suggests, constant, linear, cubic, and quintic models were rejected as they fell outside the confidence set. The model within the confidence set that was most likely to represent the data was the quartic model. This can be seen in the size of its Akaike weight, $w_i(AIC) = 0.693$, meaning that it there is a 69.3% chance that it is the best model within the set. Although the quadratic model was also in the confidence set, its Akaike weight was relatively much smaller, $w_i(AIC) = 0.252$. Thus, it would be reasonable to select the quartic function as best representing the data. Given the correspondence between this response pattern and the frequency-based training condition, we suggest that this result provides further evidence in support of our frequency-based exposure model of the uncanny valley phenomenon.

DISCUSSION

The results of Experiment 2 further qualify those obtained in Experiment 1. Although categorical performance patterns were similar to those of Experiment 1, we observed an asymmetry in categorization accuracy. Specifically, accuracy was lower for boundary-adjacent stimuli in the unequal frequency response category that was defined by fewer exemplars near the boundary, relative to the response category with more exemplars near the boundary. This suggests that the frequency training had the intended effect on categorical performance. A similar asymmetry was also observed in affective responses. However, given that measures of response times and accuracy did not correlate with affective responses, it is clear that the asymmetry in categorical perception was not the only, or even the primary, determinant of the affective asymmetry. In line with the proposed frequency-based exposure model of the UVH, we suggest that these patterns have a common cause, but that they are due to separable processes. Namely, they are both rooted in the memory traces that are encoded into long-term memory, as a result of participants' exposure to stimuli. Importantly, whereas categorical perception is the result of categorical processes which draw upon these long-term memory stores in order to determine category membership, affective responses appear to be driven by separate sub-categorical processes used to assess familiarity.

Another finding of particular importance to assessing the claims of the UVH is that the affective response pattern was in closer correspondence to the function described by Mori (1970). Specifically, we observed a slope which indicated a preference for one of the categories, and a valley region of eeriness at an intermediate region in perceptual space which was biased toward the preferred category. The primary difference between this pattern and the function described by Mori (1970) consists of the depth of the valley, and also the local affective minimum at the extreme of the preferred category. Given the similarities, we believe this pattern would provide support for a strong interpretation of UCV. These findings suggest that when a dominant reference category defined by a smaller number of high frequency exemplars is located along the same continuum as a non-dominant contrasting

category defined by a larger number of low frequency exemplars, that the reference category is associated with less negative affect. Not only does such a pattern conform to the mere exposure effect (Zajonc, 1968), but it also appears to be a reasonable generalization of the UVH proposed by Mori (1970). Thus, the critical finding of Experiment 2 is that although the pattern of responses observed in categorization performance and affective responses do not co-vary, within-category differences in exemplar frequency changed participants' affective responses. Frequency effects are thereby at least as important a determinant of the UCV as categorical perception. We consider the boarder implications of these findings below.

CONCLUDING REMARKS

In the present study, we obtained patterns across multiple dependent measures that are consistent with the UVH. Our use of affective and categorization responses further allows us to draw specific conclusions concerning the relationship between these processes, thereby establishing the phenomena of the uncanny valley as well as determining its specific properties. Our analyses of cognitive responses strongly suggests that participants perceived the stimuli categorically (Cheetham et al., 2011), as evidenced in participants' categorization responses. Our analyses of affective responses revealed patterns that were consistent with the UCV function proposed by Mori (1970), providing a weak correspondence in Experiment 1, and a strong correspondence in Experiment 2. Importantly, the lack of an association between categorical and affective responses strongly suggests that affective responses cannot be understood in terms of categorical perception. Rather, categorization responses conform to patterns that would be predicted by a category boundary in terms of categorical perception, whereas affective responses conform to patterns that would be predicted on the basis of prototype or exemplar-based models. Exemplar-based models further allow for the possibility that sub-categorical properties of the stimuli influence affective responses. Thus, in addition to providing support to the uncanny valley hypothesis, our results provide an important distinction that has not, to the best of our knowledge, been adequately made within the UCV literature. We discuss this in detail below.

UNCANNY CATEGORIES

Our observation that the uncanny valley is not solely the result of categorical perception stands in sharp contrast to recent accounts of the phenomenon (e.g., Burleigh et al., 2013; Yamada et al., 2013). In the categorization literature, there is still a lack of consensus about how category members are processed and stored. A pertinent distinction for the present discussion is whether an exemplar-based representation or a category boundary is used to classify stimuli. In contrast to early accounts that merely assumed that humans used definitions including necessary and sufficient conditions to classify stimuli, later accounts of categorization provided evidence that summary representations could play a critical role (Posner and Keele, 1968; Rosch and Mervis, 1975). A deficiency of these models, however, is that they fail to account for the retention of distributional properties of the stimuli (e.g., the distribution of all feline traits in domesticated cats) as well as particular instance (e.g., your pet cat). In the context of the present study, we cannot distinguish between category boundary and exemplar-based models of categorization performance. Difficulties in distinguishing between these models of categorization has been observed elsewhere when distributional properties have been manipulated experimentally (e.g., Stewart and Chater, 2002) as well as when these models are equated computationally (Ashby and Maddox, 1993). More specifically, difficulties in distinguishing these accounts on the basis of behavioral responses are likely a result of the general adaptability of participants to exemplar frequency in terms of category set size (e.g., Smith and Minda, 1998) and multiple learning systems (e.g., Nosofsky et al., 1994; Ashby et al., 1998). Our manipulation does, however, allow us to distinguish between alternative accounts of the UCV.

In the present study, we found strong evidence that supports the role of exemplar frequenies in determining affective responses. Unlike categorization responses wherein the category boundary was the primary determinant of performance, extrapolation items outside the initial training range were associated with greater negative affect relative to items within the training range. Such a finding is of considerable interest given that it goes against a number of well-established findings in the psychological literature. Specifically, end effects are observed when stimuli are presented along a stimulus continuum, and extreme items are identified more quickly and accurately then intermediate items (for a recent exemplar-based model, see Kent and Lamberts, 2005). Thus, when translated into the present study, we might imagine that negative affective responses should reach a minima in these regions. Similarly, these exemplars shared the smallest number of features with the contrasting category, meaning that there should be little feature mismatch. Our results indicate that a categorical perception model is inadequate in accounting for the results we obtained that support the UCV.

AFFECT AND INFORMATION PROCESSING

In contrast to categorical perception accounts of the UVH, the affective responses of our participants clearly demonstrate sensitivities to the distributional properties of categories that resulted from the manipulation of exemplar frequencies (see also, Förster et al., 2010; Gillebaart et al., 2012). Distinguishing between affective and cognitive processes should be of central importance to those interested in examining the UVH. Our results are unambiguous in differentiating between categorization performance and affective response with a sharp category boundary defining the former and graded, U-shaped (parabolic) functions defining the latter within each response category. These results might be unique to the present experimental design. A limitation of the present study is that we purposefully chose not to counterbalance the order of eeriness ratings and categorization responses. Our selection of this design followed from research that affective responses are typically produced faster than more effortful cognitive processing (Bless et al., 1990; Haidt, 2001) while decision-making appears to require that alternatives have affective valence (e.g., Damasio, 1994). A straightforward account of the present findings could be that the gradation in affective responses relative to the categorization responses was a consequence of the additional processing time resulting in activation of the category structure in long-term memory. We do not consider

this an important concern as we wished to demonstrate that frequency-based information influenced responses and acted as a determinant of the uncanny valley. Moreover, in a meta-analysis of the mere exposure effect conducted by Bornstein (1989), delays in preference ratings were shown to increase effect size. Thus, the results of the present study might reflect smaller effect sizes than are possible.

Interesting parallels have been drawn between the UVH and the speech perception literature (Moore, 2012). Namely, an examination of identification functions generally reveals strong evidence for categorical perception whereas identification response time can demonstrate slight gradations in response around the category boundary (Pisoni and Tash, 1974). Pisoni and Tash (1974) suggested that stimuli first undergo acoustic processing followed by phonemic processing. Depending on the speed of responses and the rate at which stimuli are presented, listeners can detect acoustic differences although there is a strong bias for identification on the basis of native linguistic distinctions. Additional evidence is provided by studies that have used ratings of stimulus typicality relative to a particular response category wherein listeners produce highly graded responses (Miller and Volaitis, 1989). For instance, Schoenherr and Logan (Schoenherr et al., 2012; Schoenherr and Logan, 2013, 2014) have examined individuals' performance when learning non-native phonemes wherein they were provided with feedback to reorganize a native continuum. These adult listeners appeared to be subjectively aware of the native category structure while producing identification responses that were influenced by the acoustic properties. Thus, if we were to have switched the order of affective and categorization responses then we might have observed more graded responses in the categorization response function and less graded responses in the affective response function. It is critical to note that this approach does not assess whether the sub-categorical information that is used to inform such categorization responses is affective in nature. We take the UVH to necessitate the inclusion of affective responses.

Our study is not the first to consider the role of categorical perception in the UCV phenomenon (Cheetham et al., 2011, 2014; Yamada et al., 2013). Cheetham et al. (2011) has proposed that Mori's hypothesis be considered "in terms of the well-established psychological empirical-theoretical framework of category perception and learning," and further stressed the importance of "careful definition of the category boundary" (pp. 11–12). They argued that doing so would be necessary to evaluate the potential role of categorization ambiguity in eliciting negative affect. The present study is consistent with these recommendations, and to the best of our knowledge it is the first to empirically investigate the UCV phenomenon using a category-learning paradigm.

In our study, we trained participants on stimuli belonging to non-human ontological categories to which they had little or no prior exposure. We designed our training regimen to approximate the differing levels of experience that participants would have with natural categories (e.g., human or non-human animal groups). In addition to finding evidence for categorical perception, including fitting the data with logistic functions, our analysis of affective responses demonstrated frequency effects that are not consistent with categorical perception. In a similar way to exemplar-based models that have been provided in speech perception literature to account for prototype effects (Lacerda, 1995) as well as the categorization literature more generally (Medin and Schaffer, 1978; Nosofsky, 1984), we suggest that the frequency of instances is a critical determinant of the UCV. This is consistent with accounts in the categorization literature that the frequency of training stimuli will determine the representation that is acquired by participants (e.g., Smith and Minda, 1998). Rather than seeing these results as contradictory, we suggest that sharper conceptual and methodological distinctions need to be made in terms of the contributions of affective and cognitive components. If the UCV is considered to be a product of a cognitive processes, then examinations of categorization responses are not sufficient.

The present study, as well as the current literature on the UVH, leaves open a crucial question asked by researchers studying cognition and affect: what is the causal relationship between affect and cognition? Presently, we proposed two models of the uncanny valley phenomenon. Whereas the categorical perception model assumes that categorical and affective responses are integrated at some level (Stage 3), the frequency-based exposure model assumes that they are separable (Stage 2 and Stage 3 produce different responses). Both models share in the assumption that responses are a function of stimulus comparisons with representations in long-term memory. This elementary distinction leaves open still further possibilities. Elsewhere in the affective processing literature, a number of hypotheses have been put forward which merit investigation (for a review, see Cacioppo and Gardner, 1999). Perhaps more notably, models of affect in information processing have suggested that the relationship is bi-directional, such that affect also has an influence on cognition (e.g., Bless et al., 1990; Slovic et al., 2002). In general, these models assume that affect influences the spontaneous adoption of an automatic or controlled processing strategy by signaling a benign or problematic situation (Schwarz, 2002, 2010; Schwarz and Clore, 2007). In these accounts, low-cost heuristics are relied upon when encounters are expected to go smoothly, but effortful processing is recruited when obstacles are expected. In context of the UCV phenomenon, this might suggest that uncanny stimuli increase the depth of cognitive processing. This claim is supported by studies that have an association between the amount of processing and negative affect (Bless et al., 1990). If this is the case, then we would expect to find differences in memory recall for items across an uncanny perceptual continuum, such that uncanny stimuli are more distinctive in memory (Hunt and Worthen, 2006). If studies of the UVH are to make meaningful, generalizable contributions to the literature of psychology, they must clarify how the perception of categories and affect are related. The present study represents a small step in that direction.

ACKNOWLEDGMENTS

We would like to thank Karl F. MacDorman and Mark Brown for their constructive feedback on earlier drafts of the manuscript.

STATEMENT OF ETHICS

The studies reported in this article were approved by the Research Ethics Board at the University of Guelph (REB #14JA020).

REFERENCES

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., and Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychol. Rev.* 105, 442–481. doi: 10.1037/0033-295x.105.3.442

Ashby, F. G., and Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *J. Exp. Psychol. Learn. Mem. Cogn.* 14, 33–53. doi: 10.1037/0278-7393.14.1.33

Ashby, F. G., and Maddox, W. T. (1993). Relations between prototype, exemplar and decision bound models of categorization. *J. Math. Psychol.* 37, 372–400. doi: 10.1006/jmps.1993.1023

Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. (2007). "Is the uncanny valley an uncanny cliff?," in *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium* (Jeju: IEEE), 368–373.

Bless, H., Bohner, G., Schwarz, N., and Strack, F. (1990). Mood and persuasion: a cognitive response analysis. *Pers. Soc. Psychol. Bull.* 16, 311–345. doi: 10.1177/0146167290162013

Bond, R., and Smith, P. B. (1996). Culture and conformity: a meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychol. Bull.* 119, 111–137. doi: 10.1037//0033-2909.119.1.111

Bornstein, R. F. (1989). Exposure and affect: overview and meta-analysis of research, 1968-1987. *Psychol. Bull.* 106, 265–289. doi: 10.1037//0033-2909.106.2.265

Burleigh, T. J., Schoenherr, J. R., and 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. *Comput. Hum. Behav*. 29, 759–771. doi: 10.1016/j.chb.2012.11.021

Burnham, K. P., and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach,* 2nd *Edn*. New York, NY: Springer.

Cacioppo, J. T., and Gardner, W. L. (1999). Emotion. *Annu. Rev. Psychol.* 50, 191–214.

Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., and Jancke, L. (2013). Category processing and the human likeness dimension of the uncanny valley hypothesis: eye-tracking data. *Front. Psychol.* 4:108. doi: 10.3389/fpsyg.2013.00108

Cheetham, M., Suter, P., and Jancke, L. (2014). Perceptual discrimination difficulty and familiarity in the uncanny valley: more like a "Happy Valley." *Front. Psychol.* 5:1219. doi: 10.3389/fpsyg.2014.01219

Cheetham, M., Suter, P., and Jäncke, L. (2011). The human likeness dimension of the "uncanny valley hypothesis": behavioral and functional MRI findings. *Front. Hum. Neurosci.* 5:126. doi: 10.3389/fnhum.2011.00126

Cialdini, R. B., and Goldstein, N. J. (2004). Social influence: compliance and conformity. *Annu. Rev. Psychol.* 55, 591–621. doi: 10.1146/annurev.psych.55.090902.142015

Crump, M. J., McDonnell, J. V., and Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE* 8:e57410. doi: 10.1371/journal.pone.0057410

Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY: Putnam.

Davis, F. (1930). The relative reliability of words and nonsense syllables as learning material. *J. Exp. Psychol.* XIII, 221–234. doi: 10.1037/h0071308

Fechner, G. T. (1876). *Vorschule derAesthetik.* Leipzig: Breitkoff and Hartel.

Fiske, S. T., and Taylor, S. E. (2013). *Social Cognition: From Brains to Culture.* London: Sage Publications Ltd.

Förster, J., Marguc, J., and Gillebaart, M. (2010). Novelty categorization theory. *Soc. Psychol. Pers. Compass* 4, 736–755. doi: 10.1111/j.1751-9004.2010.00289.x

Gillebaart, M., Förster, J., and Rotteveel, M. (2012). Mere exposure revisited: the influence of growth versus security cues on evaluations of novel and familiar stimuli. *J. Exp. Psychol. Gen.* 141, 699–714. doi: 10.1037/a0027612

Goldstone, R. L., Kersten, A., and Carvalho, P. F. (2012). "Concepts and Categorization," In *Handbook of Psychology,* Vol. 4, *Experimental Psychology*, *2nd Edn.,* eds I. B. Weiner, A. J. Healey, and R. W. Proctor (New York, NY: Wiley), 607–630.

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol. Rev.* 108, 814–834. doi: 10.1037/0033-295x.108.4.814

Harmon-Jones, E., and Allen, J. J. B. (2001). The role of affect in the mere exposure effect: evidence from physiological and individual differences approaches. *Pers. Soc. Psychol Bull.* 27, 889–898. doi: 10.1177/0146167201277011

Harnad, S. (1987). "Psychophysical and cognitive aspects of categorical perception: a critical overview," in *Categorical Perception: the Groundwork of Cognition* S. Harnad (New York, NY: Cambridge University Press), 1–19.

Ho, C. C., MacDorman, K. F., and Pramono, Z. D. (2008). "Human emotion and the uncanny valley: a GLM, MDS, and Isomap analysis of robot video ratings," in *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (ACM), 169–176.

Hunt, R. R., and Worthen, J. (2006). *Distinctiveness and Memory*. New York, NY: Oxford University Press.

Kent, C., and Lamberts, K. (2005). An exemplar account of the bow and set-size effects in absolute identification. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 289–305. doi: 10.1037/0278-7393.31.2.289

Lacerda, F. (1995). The perceptual-magnet effect: an emergent consequence of exemplar-based phonetic memory. *Proc. XIIIth Int. Congr. Phon. Sci.* 2, 140–147.

MacDorman, K. (2005). "Androids as an experimental apparatus: why is there an uncanny valley and can we exploit it?" in *Toward Social Mechanisms of Android Science: A CogSci 2005 Workshop* (Stresa), 106–118.

MacDorman, K. F., Green, R. D., Ho, C. C., and Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Comput. Hum. Behav.* 25, 695–710. doi: 10.1016/j.chb.2008.12.026

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

Maslow, A. H. (1937). The influence of familiarization on preference. *J. Exp. Psychol.* 21, 162–180. doi: 10.1037/h0053692

Medin, D. L., and Atran, S. (Eds.). (1999). *Folkbiology*. Cambridge: MIT Press.

Medin, D. L., and Schaffer, M. M. (1978). Context theory of classification learning. *Psychol. Rev.* 85, 207–238. doi: 10.1037/0033-295x.85.3.207

Miller, J. L., and Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Percept. Psychophys.* 46, 505–512. doi: 10.3758/bf03208147

Mitchell, W. J., Szerszen Sr, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., and MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception*, 2, 10. doi: 10.1068/i0415

Moore, R. K. (2012). A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena. *Sci. Rep.* 2, 1–5. doi: 10.1038/srep00864

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

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 104–114. doi: 10.1037//0278-7393.10.1.104

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *J. Exp. Psychol. Gen.* 115, 39–57. doi: 10.1037/0096-3445.115.1.39

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., and Glauthier, P. (1994). Comparing models of rule-based classification learning: a replication and extension of Shepard, Hovland, and Jenkins (1961). *Mem. Cogn.* 22, 352–369. doi: 10.3758/bf03200862

Paul, E. J., Boomer, J., Smith, J. D., and Ashby, F. G. (2011). Information–integration category learning and the human uncertainty response. *Mem. Cogn.* 39, 536–554. doi: 10.3758/s13421-010-0041-4

Pisoni, D. B., Aslin, R. N., Percy, A. J., and Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *J. Exp. Psychol. Hum. Percept. Perform.* 8, 297–314. doi: 10.1037//0096-1523.8.2.297

Pisoni, D. B., and Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Percept. Psychophys.* 15, 285–290. doi: 10.3758/bf03213946

Poliakoff, E., Beach, N., Best, R., Howard, T., and Gowen, E. (2013). Can looking at a hand make your skin crawl? peering into the uncanny valley for hands. *Perception* 42, 998–1000. doi: 10.1068/p7569

Posner, M. I., and Keele, S. W. (1968). On the genesis of abstract ideas. *J. Exp. Psychol.* 77, 304–363. doi: 10.1037/h0025953

Regier, T., and Kay, P. (2009). Language, thought, and color: whorf was half right. *Trends Cogn. Sci.* 13, 439–446. doi: 10.1016/j.tics.2009.07.001

Rosch, E., and Mervis, C. B. (1975). Family resemblances: studies in the internal structure of categories. *Cogn. Psychol.* 7, 573–605. doi: 10.1016/0010-0285(75)90024-9

Royall, R. M. (1997). Statistical *Evidence: A Likelihood Paradigm*. London: Chapman and Hall.

Rozin, P., and Fallon, A. E. (1987). A perspective on disgust. *Psychol. Rev.* 94, 23–41. doi: 10.1037/0033-295x.94.1.23

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., and Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* 7, 413–422. doi: 10.1093/scan/nsr025

Schneider, E., Wang, Y., and Yang, S. (2007). "Exploring the uncanny valley with Japanese video game characters," in *Proceedings of DiGRA 2007: Situated Play* (Tokyo), 546–549.

Schoenherr, J. R., and Lacroix, (2014). *Dissociating Implicit and Explicit Category Learning Systems Using Confidence Reports*. Doctoral Dissertation, Carleton University.

Schoenherr, J. R., and Burleigh, T. J. (2014). Uncanny sociocultural categories. *Front. Psychol.* 5:1456. doi: 10.3389/fpsyg.2014.01456

Schoenherr, J. R., and Logan, J. (2013). "Subjective awareness during cross-language speech perception: attending unattended regions of an acoustic continuum," in *Proceedings of the* 34th *Annual Meeting of the Cognitive Science Society* (Berlin).

Schoenherr, J. R., and Logan, J. (2014). "Attentional and immediate memory capacity limitations in the acquisition of non-native linguistic contrasts," in *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (Québec, QC)

Schoenherr, J. R., Logan, J., and Winchester, A. (2012). "Subjective confidence of acoustic and phonemic representations during speech perception," in *Proceedings of the* 34th *Annual Meeting of the Cognitive Science Society* (Sapporo).

Schwarz, N. (2002). "Situated cognition and the wisdom of feelings: cognitive tuning," in *The Wisdom in Feeling: Psychological Processes in Emotional Intelligence,* eds L. F. Barrett and P. Salovey (New York, NY: Guilford Press), 144–166.

Schwarz, N. (2010). "Feelings-as-information theory," in *Handbook of theories of social psychology,* Vol. 1, eds P. A. M. Van Lange, A. Kruglanski, and E. Tory Higgins (London: Sage Publications, Inc), 289–308.

Schwarz, N., and Clore, G. L. (2007). "Feelings and phenomenal experiences," in *Social Psychology: Handbook of Basic Principles 2nd Edn.,* eds A. Kruglanski and E. Tory Higgins (New York, NY: Guilford Press), 385–407.

Seyama, J. I., and Nagayama, R. S. (2007). The uncanny valley: effect of realism on the impression of artificial human faces. *Presence* 16, 337–351. doi: 10.1162/pres.16.4.337

Slovic, P., Finucane, M. L., Peters, E., and MacGregor, D. G. (2002). "The affect heuristic," in *Heuristics and Biases: The Psychology of Intuitive Judgment,* eds T. Gilovich, D. Griffin, and D. Kahneman (New York, NY: Cambridge University Press), 397–420.

Smith, J. D., and Minda, J. P. (1998). Prototypes in the mist: the early epochs of category learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 1411–1436. doi: 10.1037//0278-7393.24.6.1411

Steckenfinger, S. A., and Ghazanfar, A. A. (2009). Monkey visual behavior falls into the uncanny valley. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18362–18366. doi: 10.1073/pnas.0910063106

Stewart, N., and Chater, N. (2002). The effect of category variability in perceptual categorization. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 893–907. doi: 10.1037/0278-7393.28.5.893

Stroop, J. (1935). Studies of interference in serial verbal reactions. *J.Exp. Psychol.* 18, 643–662. doi: 10.1037/h0054651

Wagenmakers, E.-J., and Farrell, S. (2004). AIC model selection using Akaike weights. *Psychon. Bull. Rev.* 11, 192–196. doi: 10.3758/bf032 06482

Yamada, Y., Kawabe, T., and Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *Japan. Psychol. Res.* 55, 20–32. doi: 10.1111/j.1468-5884.2012. 00538.x

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *J. Personality and Soc. Psychol.* 9, 1–27. doi: 10.3758/bf03337428

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 10 June 2014; accepted: 03 December 2014; published online: 21 January 2015.

Citation: Burleigh TJ and Schoenherr JR (2015) A reappraisal of the uncanny valley: categorical perception or frequency-based sensitization? Front. Psychol. ***5****:1488. doi: 10.3389/fpsyg.2014.01488*

This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.

frontiers in Psychology

ORIGINAL RESEARCH
published: 30 June 2015
doi: 10.3389/fpsyg.2015.00883

Persistence of the uncanny valley: the influence of repeated interactions and a robot's attitude on its perception

Jakub A. Złotowski[1,2]*, *Hidenobu Sumioka*[2], *Shuichi Nishio*[2], *Dylan F. Glas*[3], *Christoph Bartneck*[1] and *Hiroshi Ishiguro*[2,4]

[1] Human Interface Technology Laboratory New Zealand, University of Canterbury, Christchurch, New Zealand, [2] Hiroshi Ishiguro Laboratory, Advanced Telecommunications Research Institute International, Kyoto, Japan, [3] Intelligent Robotics and Communication Laboratories, Advanced Telecommunications Research Institute International, Kyoto, Japan, [4] Department of System Innovation, Graduate School of Engineering Science, Osaka University, Osaka, Japan

OPEN ACCESS

Edited by:
Marcus Cheetham, University of Zurich, Switzerland

Reviewed by:
Julia Fink, Ecole Polytechnique Fédérale de Lausanne, Switzerland
Kurt Gray, University of North Carolina, Chapel Hill, USA

***Correspondence:**
Jakub A. Złotowski, Human Interface Technology Laboratory New Zealand, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand
jakub.zlotowski@pg.canterbury.ac.nz

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *15 June 2014*
Accepted: *15 June 2015*
Published: *30 June 2015*

Citation:
Złotowski JA, Sumioka H, Nishio S, Glas DF, Bartneck C and Ishiguro H (2015) Persistence of the uncanny valley: the influence of repeated interactions and a robot's attitude on its perception. Front. Psychol. 6:883. doi: 10.3389/fpsyg.2015.00883

The uncanny valley theory proposed by Mori has been heavily investigated in the recent years by researchers from various fields. However, the videos and images used in these studies did not permit any human interaction with the uncanny objects. Therefore, in the field of human-robot interaction it is still unclear what, if any, impact an uncanny-looking robot will have in the context of an interaction. In this paper we describe an exploratory empirical study using a live interaction paradigm that involved repeated interactions with robots that differed in embodiment and their attitude toward a human. We found that both investigated components of the uncanniness (likeability and eeriness) can be affected by an interaction with a robot. Likeability of a robot was mainly affected by its attitude and this effect was especially prominent for a machine-like robot. On the other hand, merely repeating interactions was sufficient to reduce eeriness irrespective of a robot's embodiment. As a result we urge other researchers to investigate Mori's theory in studies that involve actual human-robot interaction in order to fully understand the changing nature of this phenomenon.

Keywords: uncanny valley, anthropomorphism, human-robot interaction, multiple-interactions, eeriness, likeability, dehumanization

1. Introduction

The uncanny valley theory was originally presented by Mori (1970) in relation to a prosthetic arm. In the recent years it gathered a lot of attention in the fields of robotics, virtual agents, cognitive sciences, as well as in mass media. The uncanny valley hypothesis suggests a non-linear relationship between a robot's anthropomorphism and affinity. It proposes that by increasing humanlikeness of appearance of a robot we can also increase affinity with it. However, when a robot's appearance becomes a nearly perfect human representation, but is still distinguishable from it, people's emotional reaction instantly becomes strongly negative. Once the appearance of a robot becomes indistinguishable from a real human, the affinity with it reaches its optimum at the same level as for human beings. Furthermore, Mori suggested that movement of a prosthetic arm compared with a static arm will amplify the emotional response.

The uncanny valley is often used to explain people's rejection of anthropomorphic robots and virtual agents not only in science, but also in popular media as a reason for failure of computer-animated movies, such as The Polar Express. However, despite its wide adoption, there is relatively little empirical proof supporting it (Blow et al., 2006), e.g., the initial empirical work by Hanson (2006) and MacDorman (2006) indicated that humanlikeness might not be the only factor influencing perception of an object as eerie. Rendering style could be related with the uncanny valley for virtual agents (McDonnell et al., 2012). Moreover, it might be necessary to consider the effects of not only realism, but also the abnormality of artificial human appearance in order to investigate the uncanny valley phenomenon (Seyama and Nagayama, 2007; MacDorman et al., 2009). Mitchell et al. (2011) found that mismatch between appearance and voice can result in the uncanny valley. Furthermore, mismatch between appearance and movement of an android lead to stronger brain activation in the anterior portion of the intraparietal sulcus (Saygin et al., 2012), which could provide a neurological explanation of the uncanny valley. On the other hand, Piwek et al. (2014) reported that a realistic motion can improve acceptability especially of characters classified in the deepest point of the valley, which is against the original theory of Mori (1970) who suggested that motion will increase the uncanny effect. The uncanny valley was also reported for other primates. Monkeys looked longer at real faces and unrealistic synthetic faces than at realistic synthetic monkey faces (Steckenfinger and Ghazanfar, 2009).

1.1. Related Work

Several potential explanations have been proposed for the uncanny valley. Apart from the neurological explanation (Saygin et al., 2012), other factors included empathy (MacDorman et al., 2013), perception of experience (Gray and Wegner, 2012), threat avoidance (Mori, 1970) or terror management (MacDorman and Ishiguro, 2006). Moore (2012) provided a mathematical model using a Bayesian model of categorical perception that can explain how stimuli containing conflicting cues can give rise to a perceptual tension at category boundaries that leads to the uncanny feeling. However, studies empirically investigating categorical boundary show that ambiguous morphs close to human endpoint induce positive affect rather than negative reaction suggested by the uncanny valley hypothesis (Looser and Wheatley, 2010; Cheetham et al., 2014). Furthermore, Poliakoff et al. (2013) found that for images of prosthetic hands intermediate humanlikeness was related with the highest eeriness, but within different categories of images increased humanlikeness was related with the lowest eeriness.

Vast research efforts are also dedicated to studying the dimensions of the uncanny valley. Especially, the term used originally in Japanese by Mori (1970)— *Shinwankan*—is particularly difficult to be translated to English. Various studies used different translations, such as familiarity (MacDorman, 2006), likeability (Bartneck et al., 2009a), affinity (Mori et al., 2012), eeriness (Ho and MacDorman, 2010) or empathy (Misselhorn, 2009), which might affect the comparability of the results. Moreover, also the humanlikeness axis of Mori's graph received empirical investigation (Cheetham et al., 2011).

The shape of the graph representing the uncanny valley was disputed. In one study toy robots and humanoids were preferred even over humans (Bartneck et al., 2007). The authors proposed that the relationship between humanlikeness and likeability resembles rather a cliff than a valley, where even perfectly realistic anthropomorphic robots are liked less than toy robots or mechanoids. These results imply that building highly humanlike androids might be unfruitful as their chances of acceptance are worse than for machine-like robots. In another study Bartneck et al. (2009a) found that a highly realistic robot (android) was liked as much as a human. Furthermore, they reported that an android's realistic motion did not decrease its likeability and questioned the existence of the uncanny valley. This result is in line with a study using virtual agents (Piwek et al., 2014). However, Ho and MacDorman (2010) pointed out that the scales used by Bartneck and colleagues were correlated with warmth and as a result with each other, which might have affected the results. Overall, the literature review shows lack of agreement between different studies regarding the dimensions and the shape of the uncanny valley, and indicates that Mori's theory could be too simplistic to accurately depict the relationship between human-likeness and perception of a robot or virtual agent. Moreover, it is not clear whether this theory has any actual consequences for interaction.

1.2. Does the Uncanny Valley Affect Human-Robot Interaction?

Despite being a common research theme, the effect of the uncanny valley hypothesis on Human-Robot Interaction (HRI) is unknown. Previous studies that investigated the uncanny valley used either images or videos of different targets that were supposed to induce the uncomfortable, eerie feeling (the exception is the work of Bartneck et al. 2009a that involved short-term HRI). However, these studies did not permit any interaction between participants and robots or virtual agents. In order to understand how the uncanny valley affects HRI, it is necessary to investigate it in studies that involve physically collocated robots as their physical presence can be an important mediating factor (Kiesler et al., 2008). Previous work suggests that people's attitudes toward robots change during interaction (Fussell et al., 2008), but it has never been empirically shown whether the uncanny feeling will persist.

Little is known about the lasting effect of the uncanny valley. It is implicitly assumed that this negative emotional response toward anthropomorphic technology will have enduring consequences and lead people to reject androids that are distinguishable from humans. Since this assumption has never been verified it is important to consider an alternative hypothesis in which the uncanny valley might lead to the negative emotional response only when the target is novel and the feeling of eeriness will disappear during the course of HRI. It is possible that the affective habituation caused by repeated interactions will allow people to get used to a machine that looks almost like a human, but still is not a perfect copy. Furthermore, the uncanny valley effect might decrease when an android interacts with a human in a friendly way. If that is the case, the effects of the uncanny valley on HRI might be limited to the pre-interaction phase.

1.3. Research Questions

There is some empirical evidence suggesting only a short-term effect of the uncanny valley. In a study conducted during an ARS Electronica festival, visitors who had an opportunity to interact with an android and were interviewed afterwards, in majority, did not report an uncanny feeling (Becker-Asano et al., 2010; von der Pütten et al., 2011). Since this study had the form of an open interview that allowed people to talk freely about their experience, only a qualitative analysis was possible. Therefore, it is important to quantitatively show whether the uncanny feeling is experienced less during and after interaction with an android. Secondly, the analysis of the uncanny valley phenomenon with virtual agents indicates that there could be a relation between knowing an agent (previous exposure) and the uncanny discomfort experienced by people exposed to it (Dill et al., 2012). The decrease of previous exposure of an agent was related with higher discomfort.

Moreover, there are psychological theories that can suggest a relation between repeated exposures to a stimuli and the uncanny valley hypothesis: mere exposure effect and affective habituation. Zajonc (1968) showed that mere exposure to a neutral stimulus leads to increased positive affect toward it. On the other hand, for strongly positive or negative stimuli, the intensity of the reaction decreases after multiple exposures. This process is called affective habituation (Dijksterhuis and Smith, 2002).

The relationship between attraction and familiarity in interpersonal relations has been well documented. Positive relationships are a results of frequent face-to-face contacts (Ebbesen et al., 1976). However, if the person was disliked in the first place, greater familiarity will lead to greater dislike of that person (Ebbesen et al., 1976). This finding is consistent with work of Perlman and Oskamp (1971) who found that repeated exposure to unpleasant stimuli does not increase its likeability. Moreover, people rated more positively a person whom previously they have seen more frequently (Brockner and Swap, 1976) and they liked more others to whose ideas they were longer exposed (Brickman et al., 1975).

Four explanations have been proposed for the familiarity principle of attraction. Firstly, repeated exposure leads to increased processing fluency (Bornstein and D'Agostino, 1994), which on its own is affectively positive (Reber et al., 1998). Secondly, novel stimuli can produce uncertainty and negative reactions that diminish after a stimulus is found not to be harmful (Lee, 2001). Thirdly, due to classical conditioning, since most interactions are not aversive and rather mildly positive, others with whom people interact more often become paired with positive affect (Clark and Watson, 1988; Denrell, 2005). Fourthly, building on the previous explanation, repeated exposure creates an opportunity for interaction and these interactions are more likely to lead to rewarding social experiences (Denrell, 2005; Reis et al., 2011).

Mere exposure effect does not require interaction, but exposure is sufficient for it to occur and it has been reported for various types of stimuli (Bornstein, 1989). Although, Norton et al. (2007) proposed that in real interpersonal relations familiarity leads to dislike due to additional information about others making the less similar to oneself, Reis et al. (2011) using a live interaction paradigm showed that two previously unacquainted people shown positive affect with increased familiarity.

In relation with the uncanny valley, it is possible that for extreme stimuli the affective reaction will become weaker with people's increased familiarity with them due to affective habituation. However, for stimuli that were initially neutral, increased exposure could make them affectively more positive as a result of mere exposure effect.

This study is the first exploratory work that aims at investigating the effect of a robot's attitude and multiple interactions on the uncanny valley phenomenon by applying a live interaction paradigm in which actual HRI occurs. In particular, we focus on two aspects of interaction that could affect uncanniness of a robot: number of interactions and a robot's attitude toward a human. Moreover, we have chosen two of the most common components representing the y axis of the uncanny valley graph, *likeability* and *eeriness*, as they could be influenced differently by different aspects of HRI.

Likeability is an important factor affecting human-human relationships. Therefore, for long-term HRI it is expected to play an equally important role. There are multiple factors affecting human-human liking. One of the most important factors is history of interaction with a specific person. In particular we tend to like more others with whom we have positive rather than negative interactions (Smith and Mackie, 2007). Moreover, perception of a robot can be affected by its behavior (Goetz et al., 2003). Both positively and negatively behaving robots were anthropomorphized by people, but for an impolite behaving robot people had more mechanistic conceptions than for a positively behaving robot (Fussell et al., 2008). A robot that has a positive attitude toward a human could increase its likeability as would the classical conditioning explanation of mere exposure effect suggest. Similarly, an unfriendly robot could be liked less than it was before an interaction began. However, it is possible that an embodiment of a robot will play a role in affecting how strong effect its behavior will have on its likeability. Thus, we hypothesize that:

H_{1a}: *A friendly behaving robot's likeability will increase with repeated interactions.*

H_{1b}: *An unfriendly behaving robot's likeability will decrease with repeated interactions.*

On the other hand, we believe that previous exposure to a robot, irrespective of its behavior, will be more important for its perceived eeriness. Eerie robots could produce affective habituation and the initial strong negative emotional response will weaken with increased exposure. Similarly, for a robot that was initially perceived as neutral, repeated interactions can also positively increase the affective perception of it due to mere exposure effect.

In addition to looking at explicit measures, such as self-reports, we investigate implicit attitudes toward humanlike robots. Implicit measures assess automatic reactions and are not consciously controllable (De Houwer et al., 2009), and are incrementally valid (Steffens and Schulze König, 2006). In addition, implicit measures complement rather than replace explicit measures as they measure different aspects of the investigated attitude (Gawronski, 2002; Admoni and Scassellati,

2012). Therefore, we have also measured perceived eeriness of the robots implicitly. Thus, our next hypotheses are:

H_{2a}: Repeated interactions with a robot will reduce its explicit perceived eeriness.

H_{2b}: Repeated interactions with a robot will reduce its implicit perceived eeriness.

Recent work in HRI indicates that it might be necessary to consider anthropomorphism as a multidimensional rather than uni-dimensional phenomenon (Złotowski et al., 2014). These dimensions come from work on dehumanization—a process of depriving others of human qualities. Haslam (2006) proposed that there are two distinct senses of humanness: Human Uniqueness (HU) and Human Nature (HN). HU characteristics reflect socialization and distinguish humans from animals, e.g., intelligence, intentionality or secondary emotions. On the other hand, HN are inborn biological dispositions that distinguish humans from automata, e.g., warmth, sociability or primary emotions. Fussell et al. (2008) showed that anthropomorphism of a robot is not fixed and it changes during an interaction. It is currently unknown whether HU and HN dimensions of humanness attributed to a robot are also affected by the number of interactions or they are constant. In addition, previous work indicated that dimensions of mind attribution might be responsible for the uncanny valley phenomenon (Gray and Wegner, 2012). In particular, machines that are perceived as capable of experience, but not agency are also more uncanny. The dimensions of mind attribution and humanness are closely related (Haslam et al., 2012): agency reflects HU and experience reflects HN. Thus, our last hypotheses are:

H_3: HN, but not HU traits are related to a robot's perceived eeriness and likeability.

2. Materials and Methods

Our study was conducted using 2 ×2× 3 mixed experimental design where a robot's embodiment (humanlike vs. machine-like) and attitude (positive vs. negative) were between-subjects factors, and number of interactions (Interaction I vs. Interaction II vs. Interaction III) was a within-subjects factor. We have explicitly measured a robot's perceived eeriness, anthropomorphism, likeability, and HN and HU dimensions of humanness. Furthermore, we used the Brief Implicit Association Test (BIAT) (Sriram and Greenwald, 2009) as an implicit measurement tool of eeriness. It is a computer-based program that requires participants to classify series of words into specified categories and measures the strength of the association between these concepts and attributes using participants reaction times.

2.1. Participants

Sixty native Japanese speakers were recruited by a recruitment agency for the study. The recruitment agency for part and full-time student jobs posted on its website a message informing about the possibility of participating in a study that involves a robot. Participants were paid ¥2000 for time compensation. All participants were undergraduate students of various universities and departments located in Kansai area. Only participants who previously participated in a study involving one of the robots where excluded from selection. Due to software failure, data of two participants was corrupted or not completely saved. Therefore, we had to exclude that data from the analysis. Out of the remaining 58 participants, 26 were female and 32 were male. Their age ranged from 18 to 36 years with a mean age of 21.47. The study took place at the premises of Advanced Telecommunications Research Institute International. Adequate ethical approval was obtained from the ATR Ethics Committee and informed consent forms were signed by the participants.

2.2. Materials and Apparatus

All the implicit and explicit measurements were conducted using PsychoPy v1.78 that was run on a laptop. Participants interacted either with Geminoid HI-2 or Robovie R2. Geminoid HI-2 is the second generation of androids built as a copy of a real human (see **Figure 1**). Geminoid is indistinguishable from a human being for several seconds, until people realize its slight imperfections that lead to a negative feeling (Ishiguro, 2006; Rosenthal-von der Pütten and Krämer, 2014). On the other hand, Robovie R2 is a machine like robot that has some human features, such as a head or hands. Therefore, Geminoid HI-2 represents a robot that is near the deepest point of the uncanny valley, while humanlike features of machine looking robot—Robovie R2—should make it highly likeable (Rosenthal-von der Pütten and Krämer, 2014). Furthermore, since the uncanny valley can be also caused by a mismatch between appearance and voice or movement (e.g., Mitchell et al., 2011; Saygin et al., 2012) in order to ensure that the Geminoid HI-2 will fall into the valley we have used a synthetic child-like voice and machine-like jerky movement that does not fit the appearance of a male adult. The same movements and voice were used for Robovie R2 where the mismatch does not occur. During HRI both robots expressed idle motion that was added to increase their animacy. Geminoid HI-2 showed movement resembling blinking and breathing, as well as idle movements of its hands and synchronization of its lips to its speech. As Robovie R2 does not have a mouth, identical idle behavior was not possible. Therefore, we implemented a slight head and hand motion during speech.

FIGURE 1 | Geminoid HI-2 and a participant.

The experiment took place in a room that was divided into two parts that were separated by a folding screen in order to prevent seeing the other side (see **Figure 2**). In the experimental space a robot was placed and all HRIs occurred there. In the measurement space participants watched an introduction video that explained the order of the experiment, and they filled out all the questionnaires on a laptop. This ensured that participants did not need to judge the robot in its presence as that could have affected the results. The experimental space was equipped with cameras and the robot's behaviors were controlled by a Wizard-of-Oz who was sitting in another room.

2.3. Procedure

We used a live interaction paradigm. Participants were first shown an introduction video that explained the experimental procedure. They were told that the study involves creative and persuasive talking and they will need to convince a robot to give them a job based on the provided CV that was identical for all the participants. The experimenter ensured that participants understood the instructions and brought them to a computer. During all HRIs and filling out of questionnaires the experimenter left the participant alone in the room. The experiment was divided into 4 phases: pre-interaction video, Interaction I, Interaction II and Interaction III.

Although we have ensured that none of the participants previously interacted or participated in an experiment with the specific robot to which they were assigned, it was still possible that they have seen the robot elsewhere. In particular, in Japan it is common to see robots used in this experiment in various TV programs. Therefore, in order to minimize the differences in potential prior exposure in the pre-interaction video phase participants were asked to watch a short video (∼ 15 s) in which a robot (either Robovie R2 or Geminoid HI-2) in few sentences introduced itself and its capabilities. The dialogue was identical for both robots. After the video participants performed the BIAT and filled out all the questionnaires.

During Interaction I, participants were taken to the experimental room and sat 1.5 m in front of a robot (see **Figure 3**). They were told to have a small conversation with it to become familiar before the actual job interview begins. The robot was introduced as *Robo*. During this conversation the robot asked participants 3 neutral questions (e.g., "Is it cold today?" or "Where did you come from?"). After a short conversation was finished participants were asked to fill out the same questionnaires as the first time.

In Interaction II, the experimenter provided a short job description for which the participant was instructed to apply. Participants were asked to apply for Engineer and Bank Manager positions. The order of interviews was counterbalanced between Interaction II and III. Furthermore, a participant received a CV of a person whom she was supposed to be imitating during the interview. The CVs were identical for all participants, but the gender of applicant was always the same as the real gender of a participant. Participants were asked to use it as a base of their responses, but they could invent the information required to answer the questions. In order to motivate participants for trying to perform the task as well as they can, they were informed by the experimenter that if they secure a job, they will be paid extra money as time compensation for their participation in the experiment. They were given 5 min to prepare for the interview. After that time elapsed, the experimenter collected the CVs and job description sheets, and brought the participant to the robot.

The interview began with the robot briefly describing the company and job position for which the participant was applying. After the introduction the participant was asked 3 job interview questions. The questions were generic and common for job interviews, e.g., "Please tell me about yourself?" or "What is your biggest weakness?" While the participant was responding the robot provided feedback using non-lexical conversation sounds and non-verbal communication. In the positive condition it either nodded or nodded and uttered "Un" (expression in Japanese of agreement with the speaker). In the negative

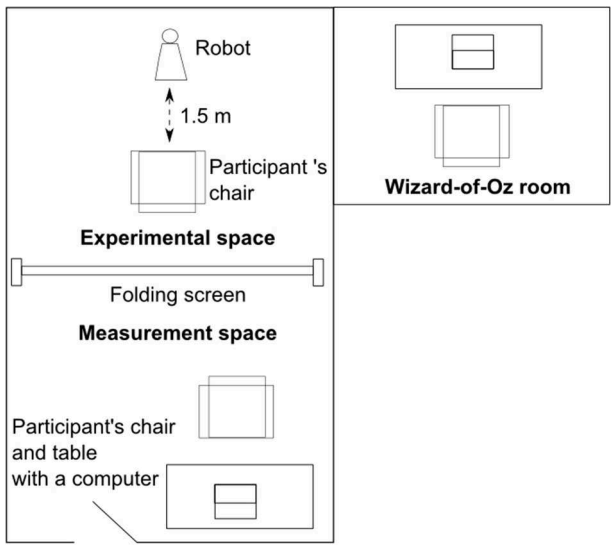


FIGURE 2 | Diagram of the experimental and measurement spaces and Wizard-of-Oz room.

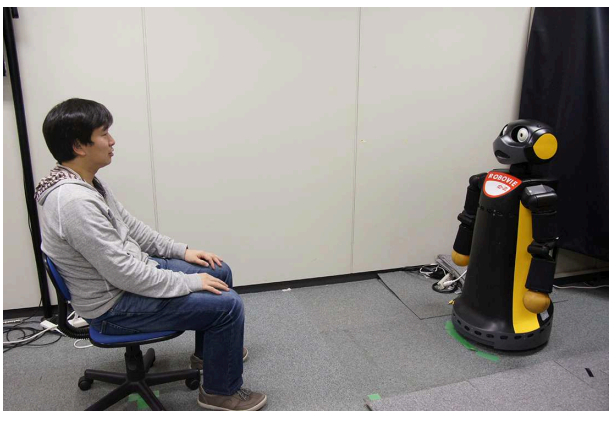


FIGURE 3 | Experimental setup. Participant sitting in front of Robovie R2 during interaction.

condition it either shook its head or nodded its head and uttered "Asso" (expression in Japanese indicating lack of interest in what the speaker says that is rather rude). This feedback was initiated by the Wizard when it was appropriate for the natural flow of conversation, e.g., when a participant paused to think about her response.

After each question the robot thanked the participant and asked the next question. After the third question the robot informed the participant that it will announce later its decision whether to give a job to a participant (in fact the decision was never announced). Although the outcome was not provided directly to a participant, the announcement varied between the conditions. In the positive condition the robot hinted approval of what the participant said during the interview. In the negative condition it was not particularly pleased with a participant's responses suggesting them to consider applying elsewhere. At that point participants were asked to fill the questionnaires for the third time. This time multiple dummy questions regarding the interview were included. Interaction III was identical as Interaction II, but the CVs, job positions and questions asked by the robot were different. Participants were permitted to answer each of the questions freely and we did not measure the duration of interactions. The whole procedure took approximately 1 h.

2.4. Measurements

In the experiment we have used several questionnaires and the BIAT (Sriram and Greenwald, 2009) as dependent measures. We explicitly measured the robots' perceived eeriness and anthropomorphism on 5-point Likert scales derived from Ho and MacDorman (2010). Moreover, likeability was measured using the corresponding Godspeed scale from Bartneck et al. (2009b) (range 1–5). In order to establish the relationship between multi-dimensional anthropomorphism and the uncanny valley we have measured 2 dimensions of anthropomorphism: HN and HU on scales developed by Haslam et al. (2009). Both dimensions had 10 items and were measured on a scale from 1 (not at all) to 7 (very much) (e.g., "The *Robo* is... shallow"). This experiment is part of a bigger study that involved additional self-report scales that were collected at the same time and are not reported here. We used a validated version of likeability scale in Japanese. Perceived eeriness, anthropomorphism, HN and HU were available only in English. Therefore, we conducted a back-translation process to obtain their Japanese versions. We calculated reliability of each scale separately for each interaction round using Cronbach's α. According to Nunnally (1978) Cronbach's $\alpha > 0.6$ is acceptable for newly developed scales for research purposes. Based on this threshold, all the scales, apart from HU were adequately reliable. The lowest Cronbach's α values during any of the three measurements were as follows: likeability $\alpha = 0.83$, perceived eeriness $\alpha = 0.62$, anthropomorphism $\alpha = 0.88$, HN $\alpha = 0.65$ and HU $\alpha = 0.54$. Low reliability of HU scale indicates that the results for this scale should be interpreted with great caution.

Furthermore, we used BIAT (Sriram and Greenwald, 2009) as a computer-based implicit measurement tool of eeriness. BIATs involve participants classifying series of words into superordinate categories. The task involved combining concept classification (*"Robo"* vs. "Human") with an attribute classification ("Eeriness" vs. "Non-eeriness"). We were interested in measuring the strength of association between *"Robo"* and "Eeriness."

In the BIAT only 2 categories are displayed on the screen at the time and in total 3 categories are being evaluated ("Interview Robot *Robo*," "Human" and "Eeriness"). The fourth category ("Non-eeriness") is called non-focal and was used only as a distractor (attribute word that does not belong to the categories that are being evaluated in a specific block) for "Eeriness." The other 2 categories ("Interview Robot *Robo*" and "Human") were used as distractors for each other. There were 2 blocks with 16 trials each that were repeated 4 times. The following stimuli were used: "Interview Robot *Robo*" (Automaton, Machine, Robot, Artificial), "Human" (Person, Natural, Mankind, Real), "Eeriness" (Eerie, Freaky, Spine-tingling, Shocking) and "Non-eeriness" (Reassuring, Numbing, Uninspiring, Boring).

At the beginning of BIAT, participants are presented with two categories that are being evaluated at the time (e.g., "Interview Robot *Robo*" and "Eeriness") and the words that belong to each of these categories. During the actual classification task these categories are displayed in the top part of the screen. At the center of the screen appear series of words that either belong to these categories or not (see **Figure 4**). Participants are asked to press as fast as possible a "K" key if the word belongs to either of the categories or "D" key if it belonged to neither category. As an example, if the categories were "Human" and "Eeriness," a participant should press "K" key if the target word is "Mankind" or "Freaky," but "D" key if the word is "Artificial" or "Reassuring." If a participant misclassified a word a red cross appeared on the screen. It remained there until the correct key was pressed.

Total time from the word appearing until the correct answer was provided was calculated with millisecond precision and was used to establish the strength of association between the categories. The idea of this task is that when an association between two categories is stronger, participants should be able to make their choices faster than for a pair of categories that are implicitly not associated with each other. The order of the BIATs was randomized and the order to blocks was counterbalanced.

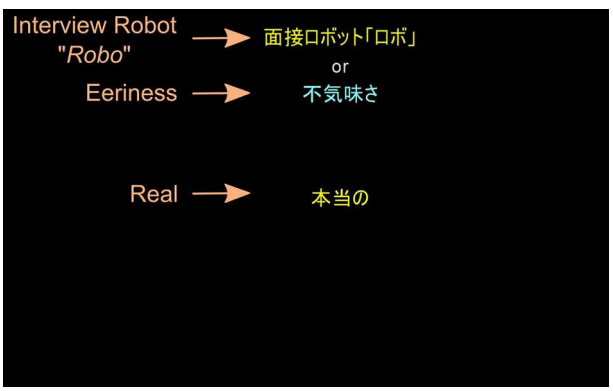


FIGURE 4 | A screenshot from the BIAT with English annotations. Two classification concepts ("Interview Robot *Robo*" and "Eeriness") and an attribute word ("Real") are being classified by a participant.

3. Results

In the first step of the analyses we looked at the explicit and implicit measures. We then looked at the relationship between these different dependent measures. To analyze the data we conducted a series of Three-Way ANOVAs with embodiment and attitude as between-subjects factors, and number of interactions as a within-subjects factor. The assumptions of used statistical tests were met, unless otherwise specified.

3.1. Likeability

First, we looked at the likeability and in particular how a robot's attitude can affect it in HRI. Due to violation of the assumption of normal distribution for parametric testing for anthropomorphism, we used a permutation test with 3 factors using the function aovp with 1000 iterations from the lmPerm package (Wheeler, 2010) using R (R Core Team, 2014). Likeability was significantly affected by the robots' attitude, $p = 0.001$ (see **Figure 5**). Positively behaving robots ($M = 3.82$, $SD = 0.67$) were liked more than negatively behaving robots ($M = 3.24$, $SD = 0.9$). Moreover, we found a statistically significant effect of embodiment with probability $p = 0.01$. Robovie R2 ($M = 3.7$, $SD = 0.88$) was liked more than Geminoid HI-2 ($M = 3.37$, $SD = 0.78$). In addition, we found a marginally significant interaction effect between embodiment and attitude, $p = 0.07$. Robovie R2 was more liked when it behaved positively ($M = 4.15$, $SD = 0.54$) than negatively ($M = 3.26$, $SD = 0.94$), $p < 0.001$. On the other hand, the attitude of Geminoid HI-2 did not significantly affect its perceived likeability.

Furthermore, we found a statistically significant interaction effect between robots' attitude and number of interactions, $p < 0.001$. During Interaction I, a robot's attitude did not affect its likeability. However, during Interaction II a robot's positive ($M = 3.86$, $SD = 0.66$) attitude increased its likeability compared to the negative attitude ($M = 2.93$, $SD = 0.98$), $p < 0.001$. Similarly, during Interaction III a robot's positive attitude ($M = 3.97$, $SD = 0.69$) resulted in higher likeability compared with a negatively behaving robot ($M = 3.2$, $SD = 0.94$), $p < 0.001$. The interaction effect between embodiment and measurement was also significant with $p < 0.001$. The difference was observed only during Interaction I when Robovie R2 ($M = 3.9$, $SD = 0.56$) was liked more than Geminoid HI-2 ($M = 3.34$, $SD = 0.61$).

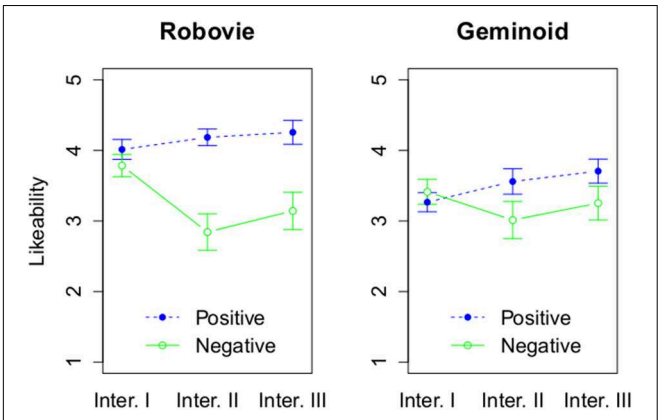


FIGURE 5 | The effect of 3 factors on likeability. The rating of likeability based on attitude and interaction round, and grouped by a robot type.

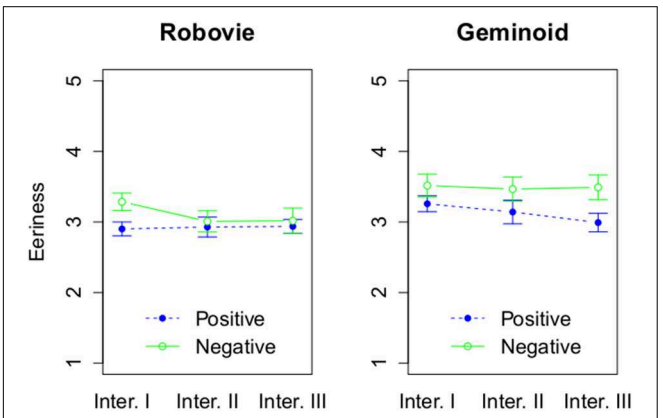


FIGURE 6 | The effect of 3 factors on explicit eeriness. The rating of explicit eeriness based on attitude and interaction round, and grouped by a robot type.

3.2. Eeriness

The second component of the uncanny valley—eeriness—was measured explicitly and implicitly. We were interested in establishing the effect of repeated interactions on a robot's perceived eeriness. Explicit measure of eeriness showed the main effect of embodiment to be statistically significant, $F_{(1, 54)} = 5.14$, $p = 0.03$, $\eta^2_G = 0.07$ (see **Figure 6**). Geminoid HI-2 ($M = 3.31$, $SD = 0.62$) was perceived as significantly more eerie than Robovie R2 ($M = 3.01$, $SD = 0.51$). Moreover, there was a significant main effect of attitude, $F_{(1, 54)} = 4.27$, $p = 0.04$, $\eta^2_G = 0.06$. A robot behaving negatively ($M = 3.3$, $SD = 0.64$) was perceived as more eerie than when it behaved positively ($M = 3.03$, $SD = 0.49$). In addition, there was a main effect of number of interactions, $F_{(2, 108)} = 3.1$, $p = 0.05$, $\eta^2_G = 0.01$. *Post-hoc* tests using the Bonferroni correction revealed that participants with marginal significance rated robots as more eerie after Interaction I ($M = 3.25$, $SD = 0.52$) than after Interaction III ($M = 3.11$, $SD = 0.6$), $p = 0.08$.

Apart from the explicit eeriness, we have also measured implicit eeriness. In the BIAT, the shorter the response time, the stronger the association between categories. The increased time would indicate that the association between a robot and eeriness is weaker. However, the reduced response time with increased number of interactions could be also due to participants improving at the task itself. Therefore, we have transformed the reaction times to z-scores within each interaction round, enabling the comparison of results between interactions. The conducted Three-Way ANOVA with embodiment and attitude as between-subjects factors, and number of interactions as a within-subjects factor did not indicate any statistically significant main or interaction effects.

3.3. Anthropomorphism

We then looked at 1 and 2-dimensional measures of anthropomorphism. We expected that there would be a main effect of a robot's embodiment and in particular Geminoid HI-2 will be perceived as more humanlike than Robovie R2. Due to violation of the assumption of normal distribution for parametric testing for anthropomorphism, we used a permutation test with 3 factors using the function aovp with 1000 iterations from the lmPerm package (Wheeler, 2010) using R (R Core Team, 2014). We found a marginally statistically significant main effect of embodiment with probability $p = 0.08$ (see **Figure 7**). Geminoid HI-2 ($M = 2.47$, $SD = 1.1$) was more anthropomorphic than Robovie R2 ($M = 2.17$, $SD = 0.92$). Moreover, we found a significant interaction effect between robots' attitude and number of interactions with probability $p < 0.001$. Only during Interaction III a robot's positive attitude ($M = 2.63$, $SD = 1.07$) resulted in higher likeability compared with a negatively behaving robot ($M = 2.11$, $SD = 1.02$), $p = 0.05$.

We then proceeded to the 2-dimensional measurement of anthropomorphism to investigate its relation with the uncanny valley. The results related to the model of anthropomorphism proposed by Złotowski et al. (2014) will be discussed in another paper. In line with previous research, we did not find statistically significant main or interaction effects for the HU dimension (see **Figure 8**).

On the other hand, we found a main effect of embodiment, $F_{(1, 54)} = 5.13$, $p = 0.03$, $\eta^2_G = 0.07$ on HN dimension (see **Figure 9**). Robovie R2 ($M = 3.16$, $SD = 0.77$) was attributed more HN traits than Geminoid HI-2 ($M = 2.74$, $SD = 0.85$). In addition, there was a significant main effect of attitude, $F_{(1, 54)} = 8.46$, $p = 0.005$, $\eta^2_G = 0.12$. Robots with positive attitude ($M = 3.21$, $SD = 0.74$) were attributed more HN than with the negative attitude ($M = 2.67$, $SD = 0.85$). There was also a significant main effect of number of interactions, $F_{(2, 108)} = 7.39$, $p = 0.001$, $\eta^2_G = 0.02$. *Post-hoc* tests using the Bonferroni correction for the family wise error revealed that the robots were attributed more HN traits after Interaction I ($M = 3.4$, $SD = 0.77$) than after Interaction II ($M = 2.88$, $SD = 0.87$), $p = 0.02$, or III ($M = 2.86$, $SD = 0.86$), $p = 0.02$. Furthermore, there was a significant interaction effect between attitude and number of interactions, $F_{(2, 108)} = 9.8$, $p < 0.001$, $\eta^2_G = 0.03$. Only for Interaction II [$F_{(1, 56)} = 15.82$, $p < 0.001$, $\eta^2_G = 0.22$] and III [$F_{(1, 56)} = 7.75$, $p = 0.007$, $\eta^2_G = 0.12$] the attitude had a significant effect.

3.4. Relationship Between the Uncanny Valley and HRI Factors

In the next step we looked at the relationship between different dependent variables used in this study in order to establish how the uncanny valley is related to factors that are important for HRI. We have calculated correlations between likeability, eeriness, 1 and 2-dimensional anthropomorphism, see **Table 1**.

The following convention was used to determine the effect size of Pearson's r coefficient: small ($0.1 \leq |r| < 0.3$), medium ($0.3 \leq |r| < 0.5$), large ($0.5 \leq |r|$). There was a correlation with large effect size between likeability and HN, $r_{(56)} = 0.54$, $p < 0.001$. Furthermore, likeability had a medium effect size correlation with

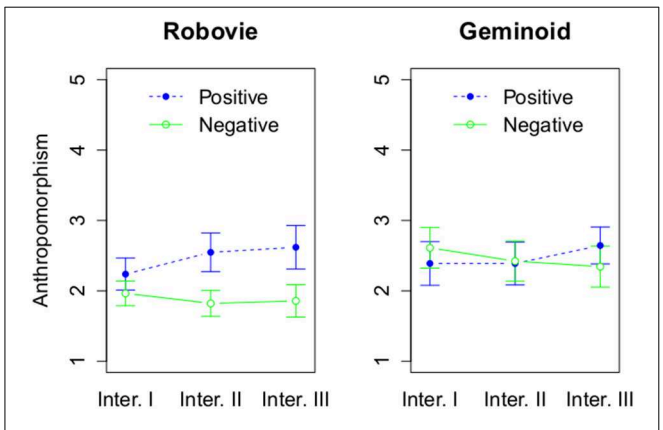


FIGURE 7 | The effect of 3 factors on anthropomorphism. The rating of anthropomorphism based on attitude and interaction round, and grouped by a robot type.

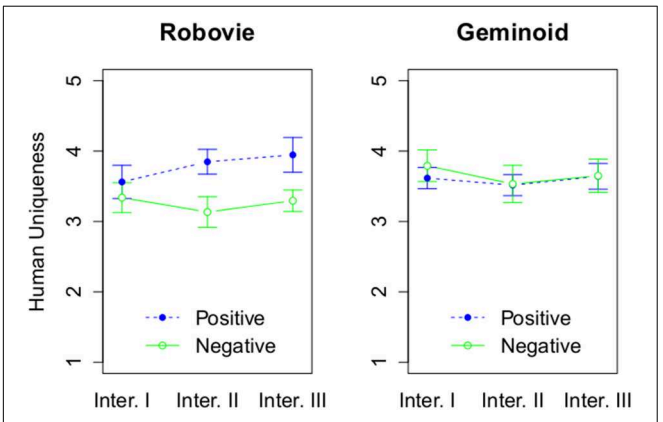


FIGURE 8 | The effect of 3 factors on Human Uniqueness. The rating of Human Uniqueness based on attitude and interaction round, and grouped by a robot type.

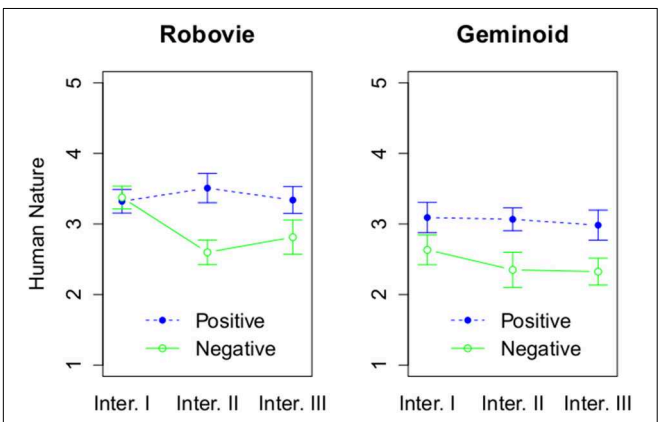


FIGURE 9 | The effect of 3 factors on Human Nature. The rating of Human Nature based on attitude and interaction round, and grouped by a robot type.

TABLE 1 | Correlations between dependent measures using Pearson's *r* coefficient.

	Likeability	Eeriness	Anthropomorphism	HU	HN
Likeability		−0.13	0.43*	0.33*	0.54*
Eeriness	−0.13		0.07	0.18	0.13
Anthropomorphism	0.43*	0.07		0.16	0.39*
HU	0.33*	0.18	0.16		0.36*
HN	0.54*	0.13	0.39*	0.36*	

*$p < 0.001$.

anthropomorphism [$r_{(56)} = 0.43$, $p < 0.001$] and HU [$r_{(56)} = 0.33$, $p < 0.001$]. Eeriness and likeability were not correlated.

4. Discussion

In this study we investigated the effect of repeated interactions and a robot's attitude on the uncanny valley phenomenon using a live interaction paradigm. In particular, we investigated the impact of these factors on a robot's likeability, as well as explicit and implicit measures of perceived eeriness. Explicit eeriness and likeability were not significantly correlated, which indicates that they measure different aspects of the uncanny valley. Although that might initially seem like an unexpected and counterintuitive finding, there are examples which show that negative correlation between eeriness and likeability is not necessary. People can dislike other people, but at the same time do not perceive them as eerie. However, there are also cases when eeriness is desirable, e.g., people who like to watch horror movies that might involve eerie creatures. Therefore, measuring both of the aspects can result in a richer picture than if we consider only one of them.

The analysis of likeability showed the more machine-like robot (Robovie R2) to be more liked than the highly humanlike Geminoid HI-2. Moreover, a robot's attitude toward a human interaction partner could be used to affect its likeability, with friendly robots being liked more than unfriendly behaving robots. However, the effect of a robot's attitude is not independent of its embodiment. The interaction effect between embodiment and attitude is especially profound in the case of a more machine-like robot. Although Robovie R2's positive behavior resulted in a small increase of likeability, it is the negative attitude that resulted in a drop of likeability ending at the level similar to the one observed for the negatively behaving Geminoid HI-2. In case of the latter robot, its attitude did not affect significantly its likeability. Thus, H_{1a} and H_{1b} are not supported.

These results seem to indicate that a robot that is perceived as uncanny is not able to affect its likeability by a positive or negative interaction. In that sense its lower likeability is persistent. On the other hand, the impact of a machine-like robot's attitude is much greater and especially when it behaves negatively, it can lose all its initial likeability. The less humanlike a robot is, the stronger that effect could be. In this study we have used only 2 robots. In **Figure 10** we present how hypothetically this relationship between humanlikeness and a robot's attitude on its likeability could look like for the broader spectrum of robots. Future, studies

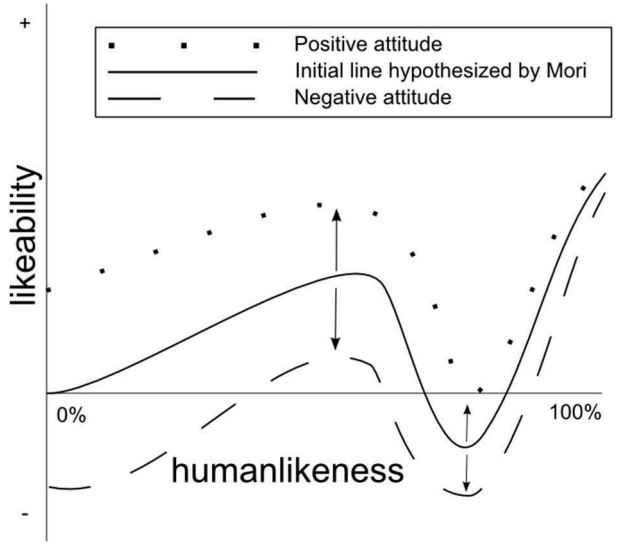


FIGURE 10 | Hypothesized effect of robots attitude on the uncanny valley. Likeability of a robot will increase with its positive attitude toward a human interaction partner or decrease with its negative attitude. The effect will be stronger the less humanlike a robot is.

are needed in order to verify how well this figure represents robots with different levels of humanlikeness than those used in this study.

These findings on likeability can also provide a new perspective on the psychological theories related with the effect of familiarity. In particular, the results are consistent rather with mere exposure effect rather than affective habituation. As suggested by the work of Perlman and Oskamp (1971); Ebbesen et al. (1976), greater familiarity with an unpleasant stimuli did not enhance liking of Geminoid HI-2, which is in contradiction with affective habituation theory. However, in case of the more neutral stimuli (Robovie R2), its behavior during interactions affected its likeability. This supports the explanation of familiarity effect proposed by Denrell (2005); Reis et al. (2011) where repeated exposure creates opportunities for interaction and those interactions that are positive due to classical conditioning will lead to a favorable impression of a person, or in this case a robot. Therefore, in live HRI mere exposure to a robot is insufficient to induce a positive affect toward it and requires a positively toned interaction. However, in case of strongly unpleasant robot, even the positive behavior can be insufficient to enhance its liking.

Looking at the second aspect of the uncanny valley investigated in this study— eeriness—we found that Geminoid HI-2 was rated as more eerie than Robovie R2. However, more interestingly we observed that after the last interaction both robots were perceived as less eerie than after interacting with them for the first time. This indicates that perceived eeriness is reduced with increased exposure to a robot. Moreover, this reduction is the same between robots that initially had different levels of eeriness, thus H_{2a} is supported. Therefore, although perceived eeriness of a highly anthropomorphic robot can decrease by merely increasing the number of HRIs, the gap

between machine-like and humanlike robots remains relatively constant. This hypothesized relationship is presented visually in **Figure 11**. Future studies involving robots with different appearances are need to evaluate the graph's exact shape.

Since both robots were perceived as less eerie after multiple interactions, it is possible that both the mere exposure effect (Zajonc, 1968) and affective habituation (Dijksterhuis and Smith, 2002) were involved in this process. Geminoid HI-2, was initially perceived as an extremely eerie robot. In this case, it is possible that affective habituation process occurred and the affective reaction became weaker with increased exposure to it. On the other hand, for an initially neutrally looking robot (Robovie R2), additional exposures were sufficient to decrease its eeriness irrespective of its behavior. Therefore, the effect of familiarity on the perceived eeriness worked differently than for likeability where a robot's positive behavior was necessary to lead to a favorable impression. If familiarity effect of attraction affects also perceived eeriness an explanation of it that requires the interaction to be positive is not supported. The more probable explanations of the obtained results for Robovie R2 are that a novel stimuli that initially fosters wary reactions after repeated interactions is found to be benign (Lee, 2001) or that additional exposures might increase a robot's processing fluency (Bornstein and D'Agostino, 1994) as its appearance becomes more familiar. Since increased processing fluency is affectively positive, it is possible that this processing affect is then transferred to the robot leading to decrease in perceived eeriness. Previous research using computer graphics investigated the relationship between the uncanny valley and these effects: exposure (Burleigh and Schoenherr, 2015), exposure and perceptual fluency (Cheetham et al., 2014), perceptual fluency (Yamada et al., 2013), and novelty and exposure effects (Cheetham et al., 2011). These experiments support our findings that repeated exposure modifies how we perceive and evaluate humanlike looking entities. Our study shows that this notion can be also applied for HRI.

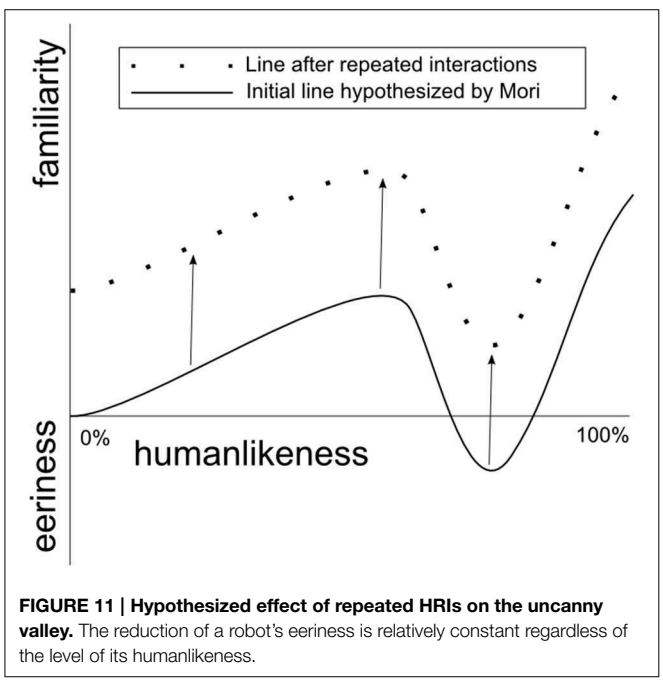


FIGURE 11 | Hypothesized effect of repeated HRIs on the uncanny valley. The reduction of a robot's eeriness is relatively constant regardless of the level of its humanlikeness.

These findings on both likeability and perceived eeriness are relevant for HRI designers. A robot can affect its likeability by its behavior. However, that effect is much stronger in case of a more machine-like robot. In particular, a machine-like robot can swiftly stop being liked despite its appearance as a result of its negative behavior. It is much harder to increase the likeability of a robot which initially falls into the uncanny valley, as a friendly attitude is not sufficient to change it.

On the other hand, people are able to quickly get used to an unfamiliar appearance of a robot. In our study three short interactions were sufficient to reduce its perceived eeriness. However, that reduction was not found to be stronger for a more anthropomorphic robot. Therefore, the relative difference in perceived eeriness between the robots remained at the same level. Nevertheless, in this study we have enhanced the eeriness of Geminoid HI-2 by creating a mismatch between its appearance, speech and movement. It is possible that if the only source of eeriness of the robot was its embodiment, the effect of multiple interactions with it would be more profound. It is also noteworthy that perceived eeriness of Geminoid HI-2 after Interaction III reached the level of Robovie R2 after Interaction I. Therefore, Geminoid HI-2 remained perceived as more eerie only because perceived eeriness of Robovie R2 also decreased. It is possible that with higher number of interactions, after a machine-like robot reaches the optimum of its familiarity, the same level can be reached by a highly humanlike robot, such as an android.

We have also found that a negatively behaving robot was rated as more eerie than a positively behaving robot. However, this finding could be explained as a result of the HRI context used in this experiment. In Japanese culture it is not typical for an interviewer to express lack of interest during a job interview in such an explicit and rude way as a robot did in this experiment. Therefore, such an attitude could have led a robot to be perceived as more eerie than when it behaved in a way that is common during human-human job interviews.

The analysis of implicit eeriness using BIAT did not show any significant differences, thus H_{2b} is not supported. Therefore, in the current form BIAT might not be optimally suited as a measurement tool of eeriness. We speculate that this result could be due to weak association between a robot's category ("Interview Robot *Robo*") that was displayed on a screen and the specific robot with which the participants interacted. Since implicit attitudes tend to change slower than explicit attitudes it is possible that our manipulation was too weak for modifying that attitude toward a specific robot. As a result, participants might have responded to the robot's category as being merely a representation of robots in general rather than their specific robotic interaction partner. In future studies, it might be beneficial to use a picture of a robot instead of a name as a representation of its category.

In line with the previous research, the HU dimension of anthropomorphism was not significantly affected by the embodiment of a robot. Furthermore, attribution of HN traits was affected by the embodiment and therefore more relevant to the uncanny valley, thus H_3 is supported. However, in contrast

with the previous work (Gray and Wegner, 2012) it was the less uncanny robot (Robovie R2) that was attributed more HN. Despite this dimension having more impact on the uncanny valley, the relationship looks to be more complex than initially proposed. The biggest difference between the work of Gray and Wegner (2012) and ours are the robots used in the experiments. In the former experiment a single robot was used that either had the back of its head visible or it had a humanlike face cover. The HN dimension is closely related with emotions and a robot that had no face is not capable of expressing emotions with facial expressions. Therefore, it was attributed less capability of experiencing (HN). In our experiment the default and fixed appearance of Robovie R2's face could be perceived as a smile. However, Geminoid HI-2 has a highly humanlike face that suggests that it can exhibit facial expressions. As a result participants might have had higher expectations, but during the interactions the robot's facial expression remained the same and was rather stern. That might have been perceived as the robot's emotional coldness and led participants to attribute less HN to it. Nevertheless, more research is needed to establish the relationship between HN and the uncanny valley. Furthermore, considering inadequately low reliability of HU dimension it is necessary to interpret these results with special care. It is possible that HU dimension is a different construct in Japan than in Western cultures.

4.1. Limitations and Future Work

In our experiment we have used only 2 robots that differed in their level of anthropomorphism. An alternative explanation for the obtained results could be that it is a robot's friendliness in appearance that is more important for its likeability than humanlikeness. We cannot exclude a possibility that there are differences along some other dimensions reflected by appearance. It is possible that if we used different pair of robots the interaction between embodiment and attitude would be reversed. In particular, Geminoid HI-2 has a stern looking facial expression, while the design of Robovie R2 could be perceived as cute and friendly with its big, childlike head. The appearance of Robovie R2 could invoke expectations for it to behave positively, and the mismatch between these expectations and the actual behavior of the robot could result in a strong decrease of its likeability. If a more friendly looking android, e.g., Geminoid F, was used in the experiment instead of Geminoid HI-2, it is possible that we would have observed a similar pattern of reactions to its unfriendly behavior as for Robovie R2. However, a question remains open why the opposite trend was not observed in case of Geminoid HI-2's mismatched positive attitude. Therefore, future studies should also include qualitative data that could help to understand why people perceive robots as eerie or likeable. Moreover, there could be demographic factors, such as age, gender or educational background, that work as moderators. The role of these factors on the uncanny valley is still not well explored.

The scale used for measuring anthropomorphism (Ho and MacDorman, 2010) in experiments of the uncanny valley was developed in a study that involved only static images of robots. However, contrary to expectations Robovie R2 and Geminoid HI-2 only marginally differed on perceived humanlikeness. Since previous work indicates that androids are perceived as more humanlike than machine like robots (e.g., Ho and MacDorman, 2010), the small difference between these 2 robots in our study must be due to other factors than merely embodiment. In order to increase the uncanniness of Geminoid HI-2 we used voice and movement that does not match its embodiment. However, the humanlikeness scale can be also affected by this manipulation as its items do not apply only to the embodiment, e.g., items rated by the participants include "Artificial"–"Lifelike" or "Fake"–"Natural." As a result our manipulation not only made Geminoid HI-2 more eerie, but also less humanlike than if only its embodiment was evaluated.

This finding also points out that a robot's behavior can be a more important factor of anthropomorphism than its embodiment. The potential solution could involve development of a new scale of anthropomorphism that is not affected by potential mismatch of a robot's embodiment and speech or movement. Alternatively, before investigating the uncanny valley in interaction it would be possible to first rate a robot's humanlikeness by presenting the static robot with no HRI.

Another limitation of this study is that participants were allowed to freely interact with a robot for as long as they wanted. Therefore, we did not consider the interaction duration in this study, but only the number of interactions. It is possible that participants who interacted with a positively-behaving robot were encouraged by its positive feedback to provide more detailed answers for their questions and as a result interacted longer with a robot. This extended interaction could have also increased familiarity of a robot and reduced its eeriness. It is also possible that the duration of interactions was insufficient to lead to the affective habituation effect of an uncanny robot. The perceived eeriness of both robots was reduced as a result of repeated interactions. However, it is still possible that after a higher number of interactions, the affective habituation effect would become stronger for the more eerie robot. A long-term study with highly anthropomorphic robots could answer this question. In particular future experiments could involve longer interactions with a robot with sessions spread over multiple days.

Future work should also consider the dynamic nature of anthropomorphism. The complexity and multifaceted nature of anthropomorphism shows a potential challenge with investigating the uncanny valley in actual, long-term HRI rather than using images or videos that can focus only on a robot's embodiment. Previous work on the uncanny valley treated it as a static feature of a robot or virtual agent. However, Fussell et al. (2008) showed that a robot's anthropomorphism changes during HRI. The results of this study also point out that at least in case of Robovie R2, its attitude affected its perceived humanlikeness. Mori's hypothesis does not accommodate for such a finding. Studies of the uncanny valley should recognize that both anthropomorphism and uncanniness of a robot can be changing during HRI, and they should consider whether the uncanny valley should be investigated using the pre-interaction level of anthropomorphism based only on a robot's appearance or the level of anthropomorphism measured in HRI at the same point of time as measures of uncanniness.

This study was an exploratory work that for the first time investigated the uncanny valley in repeated HRIs. It shows potential benefits for researching the complexity of this phenomenon in studies that involve human interaction with a collocated robot. Nevertheless, at the same time, the obtained results indicate that if we want to understand the impact of the uncanny valley on HRI, future research must go beyond picture and video based studies and enable people to interact with robots. The great majority of studies have tried to find the origin of this phenomenon. This is a worthy goal. However, until we can show that Mori's theory has any significant (long-term) impact on HRI we risk spending resources on research that might be investigating an artificial problem. In the end, it matters very little whether a picture of a robot is perceived as eerie or disliked, if during an actual interaction with a robot, this effect will vanish as a result of behavior or interaction context factors being more prominent.

Acknowledgments

This work was partially supported by Grant-in Aid for Scientific Research (S), KAKENHI (25220004) and JST CREST (Core Research of Evolutional Science and Technology) research promotion program "Creation of Human-Harmonized Information Technology for Convivial Society" Research Area. The authors would like to thank Kaiko Kuwamura, Daisuke Nakamichi, Junya Nakanishi and Kurima Sakai for their help with data collection.

References

Admoni, H., and Scassellati, B. (2012). "A multi-category theory of intention," in *Proceedings of COGSCI 2012* (Sapporo), 1266–1271.

Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. (2007). "Is the uncanny valley an uncanny cliff?," in *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication* (Jeju), 368–373.

Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. (2009a). "My robotic doppelganger - a critical look at the uncanny valley theory," in *18th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN2009* (IEEE), 269–276.

Bartneck, C., Kulic, D., Croft, E., and Zoghbi, S. (2009b). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Rob.* 1, 71–81. doi: 10.1007/s12369-008-0001-3

Becker-Asano, C., Ogawa, K., Nishio, S., and Ishiguro, H. (2010). "Exploring the uncanny valley with geminoid HI-1 in a real-world application," in *Proc. of the IADIS Int. Conf. Interfaces and Human Computer Interaction 2010, IHCI, Proc. of the IADIS Int. Conf. Game and Entertainment Technologies 2010, Part of the MCCSIS 2010* (Freiburg), 121–128.

Blow, M., Dautenhahn, K., Appleby, A., Nehaniv, C., and Lee, D. (2006). "Perception of robot smiles and dimensions for human-robot interaction design," in *The 15th IEEE International Symposium on Robot and Human Interactive Communication, 2006. ROMAN 2006* (Hatfield), 469–474.

Bornstein, R. F. (1989). Exposure and affect: overview and meta-analysis of research, 1968–1987. *Psychol. Bull.* 106, 265–289. doi: 10.1037/0033-2909.106.2.265

Bornstein, R. F., and D'Agostino, P. R. (1994). The attribution and discounting of perceptual fluency: preliminary tests of a perceptual fluency/attributional model of the mere exposure effect. *Soc. Cogn.* 12, 103–128. doi: 10.1521/soco.1994.12.2.103

Brickman, P., Meyer, P., and Fredd, S. (1975). Effects of varying exposure to another person with familiar or unfamiliar thought processes. *J. Exp. Soc. Psychol.* 11, 261–270. doi: 10.1016/S0022-1031(75)80026-6

Brockner, J., and Swap, W. C. (1976). Effects of repeated exposure and attitudinal similarity on self-disclosure and interpersonal attraction. *J. Pers. Soc. Psychol.* 33, 531–540. doi: 10.1037/0022-3514.33.5.531

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

Cheetham, M., Suter, P., and Jancke, L. (2011). The human likeness dimension of the "Uncanny valley hypothesis": behavioral and functional MRI findings. *Front. Hum. Neurosci.* 5:126. doi: 10.3389/fnhum.2011.00126

Cheetham, M., Suter, P., and Jancke, L. (2014). Perceptual discrimination difficulty and familiarity in the uncanny valley: more like a "happy valley". *Front. Psychol.* 5:1219. doi: 10.3389/fpsyg.2014.01219

Clark, L. A., and Watson, D. (1988). Mood and the mundane: relations between daily life events and self-reported mood. *J. Pers. Soc. Psychol.* 54, 296–308. doi: 10.1037/0022-3514.54.2.296

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., and Moors, A. (2009). Implicit measures: a normative analysis and review. *Psychol. Bull.* 135, 347–368. doi: 10.1037/a0014211

Denrell, J. (2005). Why most people disapprove of me: experience sampling in impression formation. *Psychol. Rev.* 112, 951–978. doi: 10.1037/0033-295X.112.4.951

Dijksterhuis, A., and Smith, P. K. (2002). Affective habituation: subliminal exposure to extreme stimuli decreases their extremity. *Emotion* 2:203. doi: 10.1037/1528-3542.2.3.203

Dill, V., Flach, L. M., Hocevar, R., Lykawka, C., Musse, S. R., and Pinho, M. S. (2012). "Evaluation of the uncanny valley in CG characters," in *12th International Conference on Intelligent Virtual Agents, IVA 2012, September 12, 2012 - September 14, 2012, LNAI*, Vol. 7502 (Santa Cruz, CA: Springer Verlag), 511–513.

Ebbesen, E. B., Kjos, G. L., and Konečni, V. J. (1976). Spatial ecology: its effects on the choice of friends and enemies. *J. Exp. Soc. Psychol.* 12, 505–518. doi: 10.1016/0022-1031(76)90030-5

Fussell, S. R., Kiesler, S., Setlock, L. D., and Yew, V. (2008). "How people anthropomorphize robots," in *HRI 2008 - Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction: Living with Robots* (Amsterdam), 145–152.

Gawronski, B. (2002). What does the implicit association test measure? a test of the convergent and discriminant validity of prejudice-related IATs. *Exp. Psychol.* 49, 171–180. doi: 10.1026/1618-3169.49.3.171

Goetz, J., Kiesler, S., and Powers, A. (2003). "Matching robot appearance and behavior to tasks to improve human-robot cooperation," in *ROMAN 2003. The 12th IEEE International Workshop on Robot and Human Interactive Communication* (Millbrae, CA), 55–60.

Gray, K., and Wegner, D. (2012). Feeling robots and human zombies: mind perception and the uncanny valley. *Cognition* 125, 125–130. doi: 10.1016/j.cognition.2012.06.007

Hanson, D. (2006). "Exploring the aesthetic range for humanoid robots," in *Proceedings of the ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science* (Vancouver, BC), 39–42.

Haslam, N. (2006). Dehumanization: an integrative review. *Pers. Soc. Psychol. Rev.* 10, 252–264. doi: 10.1207/s15327957pspr1003/4

Haslam, N., Bastian, B., Laham, S., and Loughnan, S. (2012). "Humanness, dehumanization, and moral psychology," in *The Social Psychology of Morality: Exploring the Causes of Good and Evil,* eds M. Mikulincer and P. R. Shaver (Washington, DC: American Psychological Association), 203–218.

Haslam, N., Loughnan, S., Kashima, Y., and Bain, P. (2009). Attributing and denying humanness to others. *Eur. Rev. Soc. Psychol.* 19, 55–85. doi: 10.1080/10463280801981645

Ho, C., and MacDorman, K. (2010). Revisiting the uncanny valley theory: developing and validating an alternative to the godspeed indices. *Comput. Hum. Behav.* 26, 1508–1518. doi: 10.1016/j.chb.2010.05.015

Ishiguro, H. (2006). Android science: conscious and subconscious recognition. *Connect. Sci.* 18, 319–332. doi: 10.1080/09540090600873953

Kiesler, S., Powers, A., Fussell, S. R., and Torrey, C. (2008). Anthropomorphic interactions with a robot and robot-like agent. *Soc. Cogn.* 26, 169–181. doi: 10.1521/soco.2008.26.2.169

Lee, A. Y. (2001). The mere exposure effect: an uncertainty reduction explanation revisited. *Pers. Soc. Psychol. Bull.* 27, 1255–1266. doi: 10.1177/01461672012710002

Looser, C. E., and Wheatley, T. (2010). The tipping point of animacy how, when, and where we perceive life in a face. *Psychol. Sci.* 21, 1854–1862. doi: 10.1177/0956797610388044

MacDorman, K. F. (2006). "Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: an exploration of the uncanny valley," in *ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science* (Vancouver, BC), 26–29.

MacDorman, K. F., Green, R. D., Ho, C.-C., and Koch, C. T. (2009). Too real for comfort? uncanny responses to computer generated faces. *Comput. Hum. Behav.* 25, 695–710. doi: 10.1016/j.chb.2008.12.026

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

MacDorman, K. F., Srinivas, P., and Patel, H. (2013). The uncanny valley does not interfere with level 1 visual perspective taking. *Comput. Hum. Behav.* 29, 1671–1685. doi: 10.1016/j.chb.2013.01.051

McDonnell, R., Breidt, M., and Bälthoff, H. H. (2012). Render me real?: investigating the effect of render style on the perception of animated virtual humans. *ACM Trans. Graph.* 31, 91:1–91:11. doi: 10.1145/2185520.2185587

Misselhorn, C. (2009). Empathy with inanimate objects and the uncanny valley. *Minds Machines* 19, 345–359. doi: 10.1007/s11023-009-9158-2

Mitchell, W. J., Szerszen, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., and MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception* 2, 10–12. doi: 10.1068/i0415

Moore, R. K. (2012). A bayesian explanation of the 'Uncanny valley' effect and related psychological phenomena. *Sci. Rep.* 2:864. doi: 10.1038/srep00864

Mori, M. (1970). The uncanny valley. *Energy* 7, 33–35.

Mori, M., MacDorman, K. F., and Kageki, N. (2012). The uncanny valley. *IEEE Robot. Autom. Mag.* 19, 98–100. doi: 10.1109/MRA.2012.2192811

Norton, M. I., Frost, J. H., and Ariely, D. (2007). Less is more: the lure of ambiguity, or why familiarity breeds contempt. *J. Pers. Soc. Psychol.* 92, 97–105. doi: 10.1037/0022-3514.92.1.97

Nunnally, J. (1978). *Psychometric Methods.* New York, NY: McGraw.

Perlman, D., and Oskamp, S. (1971). The effects of picture content and exposure frequency on evaluations of negroes and whites. *J. Exp. Soc. Psychol.* 7, 503–514. doi: 10.1016/0022-1031(71)90012-6

Piwek, L., McKay, L. S., and Pollick, F. E. (2014). Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition* 130, 271–277. doi: 10.1016/j.cognition.2013.11.001

Poliakoff, E., Beach, N., Best, R., Howard, T., and Gowen, E. (2013). Can looking at a hand make your skin crawl? peering into the uncanny valley for hands. *Perception* 42, 998–1000. doi: 10.1068/p7569

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reber, R., Winkielman, P., and Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychol. Sci.* 9, 45–48. doi: 10.1111/1467-9280.00008

Reis, H. T., Maniaci, M. R., Caprariello, P. A., Eastwick, P. W., and Finkel, E. J. (2011). Familiarity does indeed promote attraction in live interaction. *J. Pers. Soc. Psychol.* 101, 557–570. doi: 10.1037/a0022885

Rosenthal-von der Pütten, A. M., and Krämer, N. C. (2014). How design characteristics of robots determine evaluation and uncanny valley related responses. *Comput. Hum. Behav.* 36, 422–439. doi: 10.1016/j.chb.2014.03.066

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., and Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* 7, 413–422. doi: 10.1093/scan/nsr025

Seyama, J., and Nagayama, R. S. (2007). The uncanny valley: effect of realism on the impression of artificial human faces. *Presence* 16, 337–351. doi: 10.1162/pres.16.4.337

Smith, E. R., and Mackie, D. M. (2007). *Social Psychology, 3rd Edn.* New York, NY: Psychology Press.

Sriram, N., and Greenwald, A. G. (2009). The brief implicit association test. *Exp. Psychol.* 56, 283–294. doi: 10.1027/1618-3169.56.4.283

Steckenfinger, S. A., and Ghazanfar, A. A. (2009). Monkey visual behavior falls into the uncanny valley. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18362–18366. doi: 10.1073/pnas.0910063106

Steffens, M. C., and Schulze König, S. (2006). Predicting spontaneous big five behavior with implicit association tests. *Eur. J. Psychol. Assess.* 22, 13–20. doi: 10.1027/1015-5759.22.1.13

von der Pütten, A. M., Krämer, N. C., Becker-Asano, C., and Ishiguro, H. (2011). "An android in the field," in *Proceedings of the 6th International Conference on Human-Robot Interaction, HRI '11* (New York, NY: ACM), 283–284.

Wheeler, B. (2010). *lmPerm: Permutation Tests for Linear Models*. R package version 1.1-2.

Yamada, Y., Kawabe, T., and Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *Jpn. Psychol. Res.* 55, 20–32. doi: 10.1111/j.1468-5884.2012.00538.x

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *J. Pers. Soc. Psychol.* 9(2 Pt 2), 1–27. doi: 10.1037/h0025848

Złotowski, J., Strasser, E., and Bartneck, C. (2014). "Dimensions of anthropomorphism: from humanness to humanlikeness," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction, HRI '14* (New York, NY: ACM), 66–73.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 27 February 2015
doi: 10.3389/fpsyg.2015.00199

Perception of gait patterns that deviate from normal and symmetric biped locomotion

Ismet Handžić and Kyle B. Reed*

REED Lab, Department of Mechanical Engineering, University of South Florida, Tampa, FL, USA

Edited by:
Marcus Cheetham, University of Zurich, Switzerland

Reviewed by:
Burcu Aysen Urgen, University of California, San Diego, USA
Matthieu Destephe, Waseda University, Japan
Shuichi Nishio, Advanced Telecommunications Research Institute International, Japan

***Correspondence:**
Kyle B. Reed, Department of Mechanical Engineering, University of South Florida, 4202. E. Fowler Ave., ENB118, Tampa, FL 33612, USA
e-mail: kylereed@usf.edu

This study examines the range of gait patterns that are perceived as healthy and human-like with the goal of understanding how much asymmetry is allowable in a gait pattern before other people start to notice a gait impairment. Specifically, this study explores if certain abnormal walking patterns can be dismissed as unimpaired or not uncanny. Altering gait biomechanics is generally done in the fields of prosthetics and rehabilitation, however the perception of gait is often neglected. Although a certain gait can be functional, it may not be considered as normal by observers. On the other hand, an abnormally perceived gait may be more practical or necessary in some situations, such as limping after an injury or stroke and when wearing a prosthesis. This research will help to find the balance between the form and function of gait. Gait patterns are synthetically created using a passive dynamic walker (PDW) model that allows gait patterns to be systematically changed without the confounding influence from human sensorimotor feedback during walking. This standardized method allows the perception of specific changes in gait to be studied. The PDW model was used to produce walking patterns that showed a degree of abnormality in gait cadence, knee height, step length, and swing time created by changing the foot roll-over-shape, knee damping, knee location, and leg masses. The gait patterns were shown to participants who rated them according to separate scales of impairment and uncanniness. The results indicate that some pathological and asymmetric gait patterns are perceived as unimpaired and normal. Step time and step length asymmetries less than 5%, small knee location differences, and gait cadence changes of 25% do not result in a change in perception. The results also show that the parameters of a pathologically or uncanny perceived gait can be beneficially altered by increasing other independent parameters, in some sense masking the initial pathology.

Keywords: uncanny valley, gait perception, biped walking, passive dynamic walking, gait simulation, pathological gait, rehabilitation

1. INTRODUCTION

In order to systematically generate a variety of altered gait dynamics to be rated for impairment and uncanniness by participants, we are using one measured healthy gait (for comparison) and sets of simulated gait models that are mathematically derived. Simulated gait models allow for consistency and precision of the altered gait parameters. This systematic change allows for a controlled experiment on the perception of specific gait changes. By using a passive dynamic walker (PDW) computational model, we are able to specifically examine changes in perception that arise from deviations in gait speed, knee location, spatial and temporal symmetry, foot roll-over shapes, and knee damping.

A healthy human body with a human-like shape and movements is perceived as normal, healthy, and familiar. Also, an exaggerated caricature of a human body and its animated movements can be accepted as somewhat normal and familiar as we expect the caricature to be un-human-like. However, human-like objects, models, robots, or dolls often are designed to mimic normal human body parts, motions, or gestures that almost look normal, but cause an eerie feeling. This psychological reaction to the almost human-like is known as the uncanny valley (Jentsch, 1906; Freud, 1919; Mori, 1970; Eberle, 2012). The uncanny valley can sometimes be described as the perception of something that is familiar, yet incongruous, creating a repulsive effect.

Although the notion of the uncanny valley is widely known, the depths and edges of it are still fuzzy and open for study. It is not clear what changes from normal and human-like will cause one to perceive the altered motions with feelings of uneasiness. As shown in **Figure 1**, the initial proposal of the uncanny valley is defined as the descent of the plot between human likeness (horizontal axis) and our familiarity (vertical axis) (Mori, 1970).

This decent and the relationship between familiarity and human likeness was initially shown to vary to where moving human-like objects fall further into the uncanny valley than still objects. However, a recent study that examined the existence of the uncanny valley for still and moving human-like objects concluded that, opposed to static human-like characters, augmented human walking movements will not cause any dip in

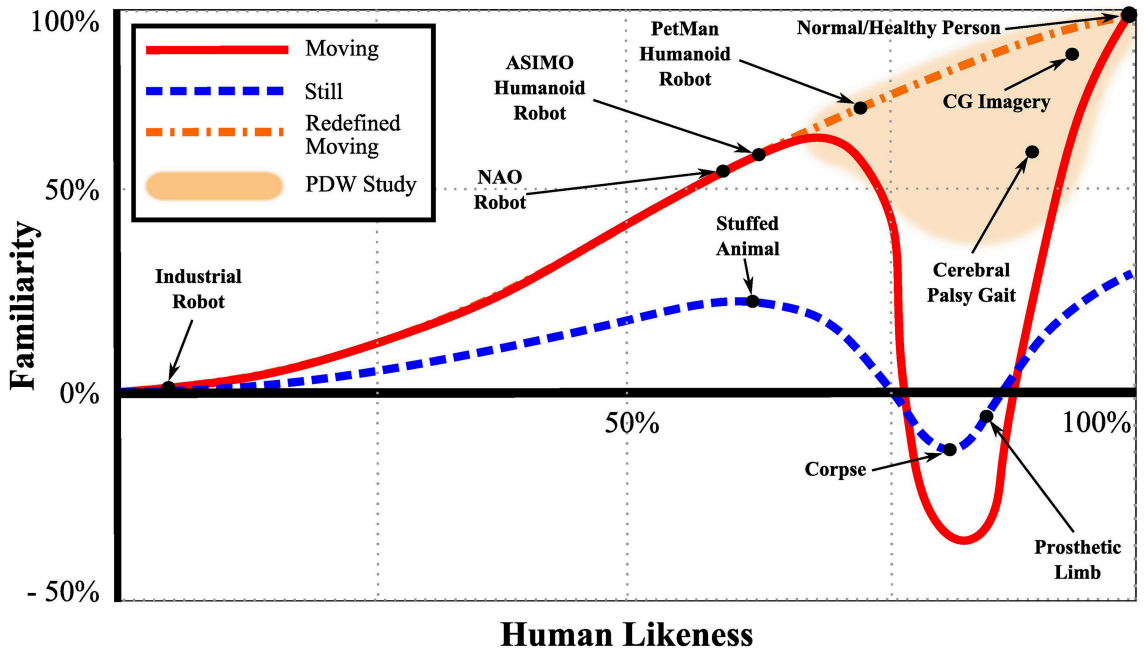


FIGURE 1 | The Uncanny Valley of still and moving objects. For still objects, as human likeness increases, so does the observer's familiarity up until a sudden decrease in familiarity into the uncanny valley where familiarity becomes abnormal, perceived with an uneasy or eerie feeling. Initially Mori (1970) proposed the uncanny valley for moving objects, however no study was found to reinforce the concept of the uncanny valley for moving objects. Recent studies with synthetic agents (Thompson et al., 2011; Piwek et al., 2014) show proof that moving objects tend not to fall into the uncanny valley but rather monotonically increase in familiarity with human likeness. The focus of the study presented here is on the right side that relates to what deviations from normal and human-like cause the perception to decrease.

familiarity with increased human likeness. That is, the relationship between familiarity and human-likeness changes monotonically with augmented walking (Piwek et al., 2014). Furthermore, another study by Thompson et al. indicated similar results deducing that when human walking motion parameter changes (joint dis-articulation, jerk, and phase movement changes) are examined, the familiarity rating of a synthetic agent (augmented human motion computer graphic character) by human observers do not show the uncanny valley (Thompson et al., 2011). Although there are studies that verify the uncanny valley for human faces (Seyama and Nagayama, 2007; MacDorman et al., 2009), we were not able not find a clear study that proves the uncanny valley for human body motions. It is interesting to note that one study showed an improvement in familiarity with human likeness in faces with motion compared to still faces (McDonnell et al., 2012).

As we approach the familiarity vs. human-likeness function from the left (low human likeness), we encounter it with lifeless objects, models, and movements such as industrial robots, stuffed puppets, or humanoid robots. The left side of the valley is characterized by motions and attributes that we know not to be human, but have some characteristics that are human-like. However, approaching this function from the right (high human likeness), that is, coming from the perception of a normal and healthy person, the body motions are highly realistic and match our expectation of how a normal and healthy human typically moves. The top-right side of the valley is populated by very human-like features and motions, however may show some traits that are not exactly normal or healthy. In this article we focus on the right side which is shaded in **Figure 1**. Specifically, we examine the perception of human walking motions and the limits to which gait will continue to be perceived as normal and human-like in the presence of abnormalities.

Our hypothesis is that gait can appear human-like even when it deviates from perfect temporal and spatial symmetry. Although there are distinct kinematic differences between walking in tennis shoes and high-heeled shoes (Hansen and Childress, 2004), both exhibit a healthy familiar human-like gait. Contrarily, walking with a badly sprained ankle is quickly noticed as a limping gait. Uncanniness emerges when a motion or appearance is close, but not exactly as expected, similar to the feelings that arise when one views individuals walking with a severe injury or disability (Lipson and Rogers, 2000; Henderson and Bryan, 2004). The focus of this study is on the motions that constitute the gait and how to reduce the perception that a gait pattern is abnormal. These results could guide physical therapists in their treatments and would benefit individuals with disabilities that affect gait by determining the gait patterns that minimize the perception that their gait is impaired. Appearance is a major concern for individuals with a disability (Bohannon et al., 1988, 1991). That is, an individual may have the functional ability to walk and it is important for them to be perceived as normal as possible.

2. BACKGROUND

2.1. HUMAN GAIT

To ensure understanding of the gait deviations described throughout this paper, we will provide a short background on normal and impaired gait patterns. Normal walking in healthy and unimpaired individuals is smooth and combines complex balancing, shock absorbing, and propelling dynamics along with central nervous system signals to generate efficient locomotion. In a healthy gait pattern, both legs move symmetrically and mirror all dynamics 180° out of phase. As opposed to running, individuals retain ground contact throughout the gait cycle (Perry, 2010; Whittle, 2012). The repeating gait cycle can be subdivided into two periods (stance and swing), eight phases (heel strike, loading response, mid stance, terminal stance, toe-off, initial swing, mid swing, and terminal swing), or three tasks (weight acceptance, single limb support, and limb swing) (Perry, 2010; Whittle, 2012). Some of these subdivisions of normal gait can be seen in **Figure 2**. The upper body, which includes head, neck, trunk, and arms, moves along as a unit and is considered the passenger unit to the locomotor system, which consists of the legs (Perry, 2010).

Normal healthy walking is symmetric in left-right step length distance, leg swing time, internal joint forces, and external ground reaction forces. The concept of gait symmetry in able-bodied human beings is still an on-going debate (Sadeghi et al., 2000). While many studies exist that assume gait symmetry for the sake of simplicity in data collection analysis, other studies assume gait symmetries if no statistically significant differences are noted on parameters (kinematics or kinetics) measured between limbs. Most able-bodied individuals inherently have some small and unnoticeable spatial and temporal gait asymmetries due to limb dominance or frequent and demanding movements such as in sports (Sadeghi et al., 2000).

An important aspect of gait is the roll-over shape (ROS) that the foot effectively follows when completing the stance phase during the gait cycle. ROSs of a healthy person during stance phase is presented in **Figure 3**. ROS have enormous effects on gait kinematics, kinetics, and balance (Menant et al., 2009), and ROS are important in prosthetic design (Hansen et al., 2000; Curtze et al., 2009; Hansen and Wang, 2010). The forces exerted on a foot or by a prosthetic leg onto an individual can be manipulated if the ROS is modified properly (Rietman et al., 2002).

Gait pathology can come in various forms such as deformity, muscle weakness, sensory loss, pain, and impaired motor control caused by disease, injury, or genetic birth traits (Perry, 2010).

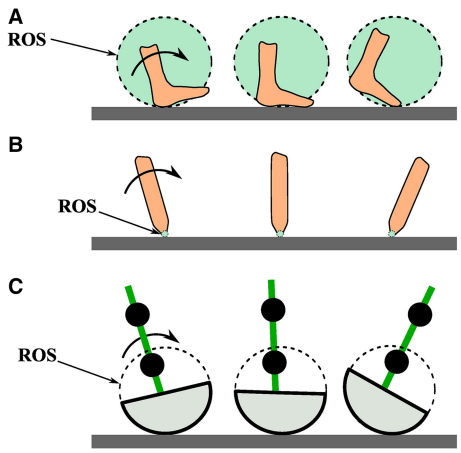


FIGURE 3 | (A) A person's foot rolling over the path of its roll-over shape (ROS). **(B)** A point contact rolling over its roll-over shape. **(C)** This study's passive dynamic walker foot rolling over its roll-over shape.

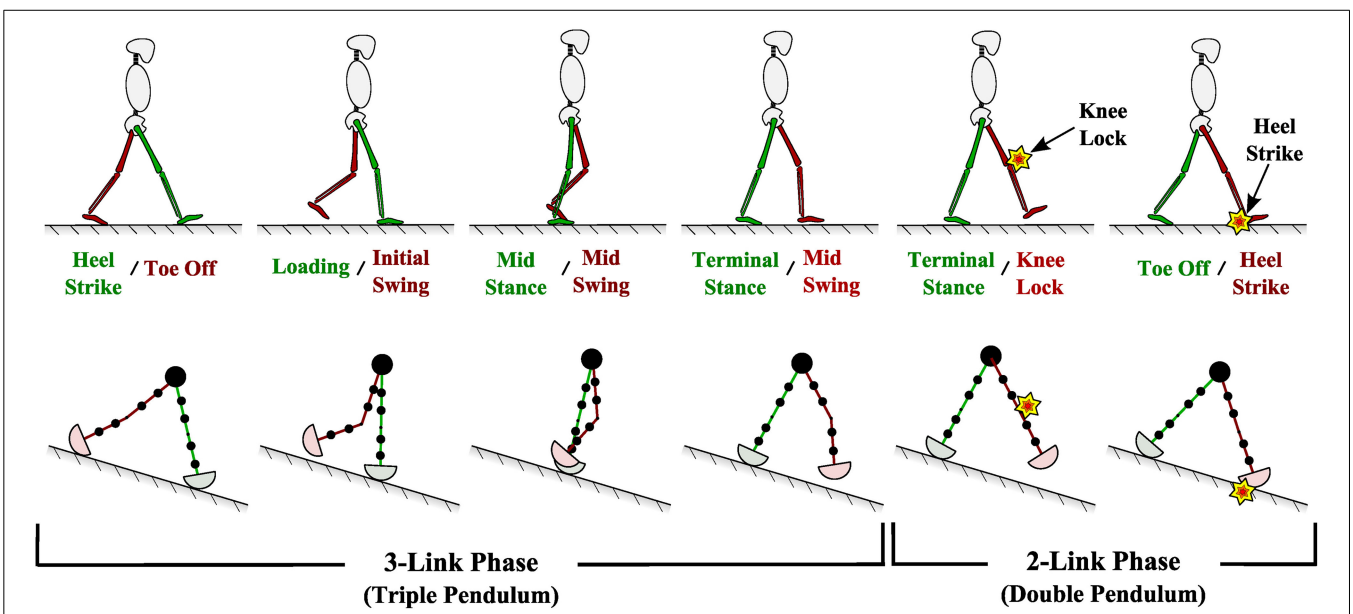


FIGURE 2 | (Top) Phases of normal and healthy human walking. (Bottom) Passive dynamic walking computational model used in this study to generate various walking patterns.

Such gait pathologies can cause mild or severe gait dynamics or ROS deviations, which may or may not be easily recognizable by other individuals. Deviations from normal walking is often accompanied by compensatory leg dynamics, which may be damaging to other parts of the body. For instance, a person wearing a leg prosthetic which is geometrically identical to the opposite healthy limb, may exhibit recognizable compensatory dynamics such as asymmetric step length, swing time, internal force, or foot ROS asymmetries (Schmalz et al., 2002; Rabufetti et al., 2005; Curtze et al., 2009). ROS for healthy humans can be approximated to be of constant radius and one-third the length of the leg (McGeer, 1990; Adamczyk et al., 2006).

2.2. PERCEPTION OF GAIT

Humans are very effective at recognizing other humans and the complex motions exerted by other humans (Kozlowski and Cutting, 1977; Loula et al., 2005). While this perception has been generally studied (Blake and Shiffrar, 2007), the creating of the gait perception stimuli has varied. By walking on an asymmetric split-belt treadmill, it has also been shown that humans are able to recognize gait asymmetry in their own gait when walking asymmetry exceeded a specific threshold (Lauzière et al., 2014). The gait parameter that corresponded the most with belt speed asymmetry was found to be stance time.

While other forms of methods to recreate human motion for perception analysis has been studied in the past such as PL animation of biological motion (Lee et al., 2002), motion capture (Knoblich and Flach, 2001), or morphing of bipedal locomotion movements (Giese and Lappe, 2002), no study exists that uses a purely dynamics model to evaluate the perception of human gait by systematically altering such a dynamics gait model. This study aims to use a modeled biped model as a perception stimuli by systematically altering the model's dynamics by manipulating its parameters.

2.3. UNCANNY VALLEY AND PATHOLOGICAL GAIT

Humans are keenly aware of walking motions that are close, but not exactly the same as a human makes. To other human observers, a normal healthy gait does not draw any attention and is usually dismissed as ordinary. However, as normal and healthy walking becomes unhealthy or impaired, it starts to raise attention and sometimes uneasy feelings, hence sometimes raising uncanny (eerie) feelings toward the gait mechanics. At an extreme end, this uncanny feeling can be provoked when observing the gait of extremely walking-impaired individuals suffering from neurological movement disorders such as athetoid cerebral palsy or dystonia, resulting in involuntary muscle contractions, repetitive movements, or abnormal postures. However, even smaller alterations from normal healthy gait may be easily recognizable and viewed as abnormal or unfamiliar. Pathological human gait, such as a slightly limping leg or sprained ankle, can be viewed as human-like and normal, yet the impairment will be quickly identified.

In healthy humans, the two sides of the body are mostly symmetric with regards to mass and strength; thus, it makes biomechanical sense to have both knees at the same location. However, when wearing a transfemoral prosthesis, the mass and strength of the two legs are no longer equal and the biomechanical reasons to keep the same prosthetic knee location no longer exist. Moving the knee location adds a degree of freedom in the prosthesis design process that allows the gait dynamics to be adjusted to a desired gait pattern (Sushko et al., 2012). However, changing the knee location depends on the answer to an essential question for this study: what amount of knee location asymmetry can be considered normal or human-like? Note that we are only concerned with the bio-mechanical movements of leg limbs and how these movements are perceived in this study. We are not investigating the effects of limb thickness or texture perception, such as wearing a Flex-Foot Cheetah prosthetic blade foot (Grabowski et al., 2010).

2.4. UNCANNY VALLEY AND ARTIFICIAL GAIT

Toyota's ASIMO (Sakagami et al., 2002) and Aldebaran Robotics's NAO (Anderson et al., 2011) robots are statically stable robots that are able to simulate a slow and careful walking pattern while always keeping their center of gravity above their support base. Humans can walk this way, but rarely do. Such statically stable robotic gait is only partially perceived as human-like and can come off as stiff, "robotic," and sometimes uncanny. While more proficient in its gait, Boston Dynamics's PetMan (Raibert, 2010) is an anthropomorphically correct biped able to mimic gait very similar to humans. PetMan is able to skillfully navigate across obstacles such as stairs and withstand moderate perturbations during gait. Nonetheless, its more realistic motions invoke an unhuman-like perception of its movements. These humanoid robots are perceived to be on the left side of the uncanny valley and so are of little direct interest to our study and hypothesis about the right side of the valley.

On the other hand, dynamically stable walking robots such as a passive dynamic walker (PDW), exhibit a more fluent and human-like gait. A PDW is a biped walking robot that walks down a decline with gravitational energy as its only source of power and with no active feedback (McGeer, 1990). PDW gait is shown to be kinematically and kinetically similar to human gait (Adamczyk et al., 2006; Kuo, 2007; Handžić and Reed, 2013a,b). While PDWs can be used to recreate and analyze normal and pathological human walking patterns, they can also be utilized to study the effects on gait caused by manipulating swinging limb parameters such as leg lengths, leg masses, joint stiffness, or ROS (Honeycutt et al., 2011).

3. MATERIALS AND METHODS

3.1. PASSIVE DYNAMIC WALKING GAIT

This study employed a PDW computational model because the PDW model is repeatable, precise, and can be systematically altered in order to implement altered gait patterns. This consistency allows the controlled variation of desired parameters (i.e., step length, limb mass, joint stiffness, ROS etc.) without the inconsistency of human sensorimotor control under the same walking conditions.

The PDW model is a two dimensional nine-mass multi-pendulum system with constant-radius-shaped feet. That is, it represents an anthropomorphically correct walking human from the waist down and viewed from a two dimensional sagital

plane. PDW masses are represented as one hip mass and two masses per each thigh and shank. The PDW model also rolls over a constant radius roll-over shape just as a walking human would (**Figure 3C**). Just as in human gait, the PDW legs progress through two distinct phases, stance and swing, as it advances down a decline as seen in **Figure 2**. During a step and before knee lock, the PDW is modeled as an inverted triple pendulum as the shank swings forward, after which it turns into an inverted double pendulum. The kinematics of our PDW can be derived with the Lagrangian formulation, while the knee lock and heel strike collision events can be described with conservation of angular momentum. The mathematical modeling for our PDW with point feet can be reviewed in McGeer (1990), Chen (2005), and Honeycutt et al. (2011). Although, the PDW can walk down a greater decline, our model walks down a slope of 3.5° for all gait variations presented in this study. We specified the PDW model height, thigh length, and shank length, mass and mass distribution according to widely surveyed anthropomorphic body segment data (Drillis et al., 1964). The roll-over shape for normal walking was taken to be one-third leg length as found in Adamczyk et al. (2006). All PDW deviations presented in this study were stable for at least fifty steps.

3.2. MEASURED NORMAL GAIT

In addition to the systematically altered gait patterns derived from the PDW modeled gait, one gait pattern was collected from a healthy individual walking at a comfortable speed over level ground. The individual was 28 years of age, 93 kg (205 lb), and was 1.85 m (6 ft 1 in) tall. The individual walked barefoot on a stationary treadmill at 0.8 m/s. The treadmill and the motion tracker system are part of the Computer Assisted Rehabilitation Environment (CAREN) system. The gait was recorded using a VICON® motion capture system with ten Bonita B10 cameras set to record at 100 Hertz. Reflective markers (14 mm in diameter) were placed on both left and right hip (anterior superior iliac spine), knee joint, ankle joint, and big toe (phalanges). The individual walked for ten strides at steady state and an average of those motions was used as the comparison video. The individual whose gait was recorded signed a University of South Florida Institutional Review Board (IRB) consent form before volunteering to be analyzed for this study.

3.3. PASSIVE DYNAMIC WALKING ANIMATION VIDEOS

Because this study predominantly focuses on normal and abnormal human walking motions, the PDW model closely depicts the aesthetics of a person walking when viewed from the side (silhouette). This helps to increase the participant's familiarity and human likeness of the presented walking models. The animation silhouette was closely depicted to mimic human muscles, joints, knees, and feet by considering waist, mid-thigh, and max calf circumference as outlined by the United States Department of Health and Human Services Health Statistics Report (McDowell et al., 2008). This aesthetic transformation of our PDW model can be seen in **Figure 4**. Note that the focus of this study is on the motions of the gait and not the static appearance of the legs. Although the PDW walks down a decline, it was rotated to look as if it is walking on level ground. Feet were animated by interpolating the foot angle trajectory of the actual recorded normal gait and fitting it onto the computational dynamics of the PDW model since the PDW model does not simulate feet.

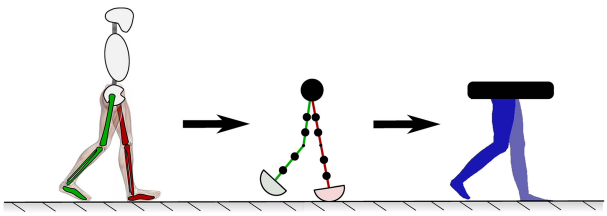


FIGURE 4 | The human lower limb model presented to participants was modeled with a PDW to move naturally while being carefully depicted to look like normal human limbs.

While previous uncanny valley studies that analyzed body motions used computer generated animations controlled by deteriorated and augmented human body motions (Piwek et al., 2014; Thompson et al., 2011), the gait motions in this study are based on computational dynamics allowing systematic and precise augmentation or leg kinematics.

Various PDW walking parameters were computationally and systematically varied to deliberately deviate from familiar and human-like (normal) gait to explore the perception of impairment and uncanniness of human gait. Although the PDW computational model can simulate many parameters with any parameter resolution, that would yield many videos to be judged by participants, which would result in a prolonged experiment per participant. **Table 1** shows the different parameter categories chosen for this study that were presented to the participants. All PDW leg variations were applied to the leg closest to the observer (i.e., darker, right). Note that Equation (1) is used to define percent asymmetry between two parameters. Equation (1) is used to define percent asymmetry for all parameters. The negative values in **Table 1** for knee height refer to a decrease in knee location and the negative values in gait cadence refer to a slower speed.

$$Asymmetry\ (\%) = \left(\frac{abs(Left - Right)}{(Left + Right)/2} \right) \tag{1}$$

3.4. GAIT VIDEOS

The following videos were presented to participants. Participants judged all videos on the the basis of two separate metrics: impairment and uncanniness.

3.4.1. Measured normal gait

This gait pattern was recorded from a healthy individual who had no asymmetries or abnormalities. The recorded gait cadence was measured at 80 steps/min. This video was included to compare the perception of the PDW modeled normal gait to a human gait. Note that the measured normal gait from this human participant has an approximately 25% slower cadence than the modeled normal gait. This difference highlights the benefit of the PDW model for allowing a systematic alteration of the gait patterns; we cannot impose a specific change in a human, but can in the PDW modeled system.

Table 1 | Five PDW parameter categories were studied.

Gait cadence (%)	Knee height asymmetry (%)	Spatial and temporal asymmetry (%)	ROS asymmetry (%)	Knee damping with mass asymmetry (%)
−50	+83	5 (LaTa)	29	0
−25	+57	13 (LsTa)	66	40
+25	+22	5 (LaTs)	100	100
+50	−26	13 (LaTs)		118
	−40	5 (LaTa)		
	−61	13 (LaTa)		

The 23 listed here, plus recorded and PDW modeled normal videos were presented. L = Step Length, T = Swing Time, s = Symmetry, a = Asymmetry.

3.4.2. PDW modeled normal gait

The normal PDW modeled walking pattern was perfectly symmetric between left and right sides. This normal gait walking cadence was matched to that of a healthy adult walking cadence at 110 steps/min (Perry, 2010). This video was shown as a baseline and for comparison to a recorded walking pattern from a healthy human participant. This video was also used as the stimulus (base) for comparison in each category.

3.4.3. Category 1: Gait cadence

Gait cadence may affect the observer's perception of the gait, so four different videos of the PDW modeled normal gait at four different speeds were included in the study (two slower and two faster) (−50, −25, +25, and +50%).

3.4.4. Category 2: Knee height

As previously reviewed in the background section, prosthetic knee location (knee height) may be altered in order to gain spatial, temporal, kinetic symmetry, or comfort while walking. These alterations aim to determine how much deviation in knee height symmetry is noticeable and perceived as uncanny. As listed in **Table 1**, we present three videos where the walking model has a knee asymmetry with one knee raised and three videos that show the walking model with knee asymmetry by lowering one knee. All models in this category have symmetric step lengths and swing times. Because the knee is displaced very close to the hip, the video with +83% knee height shows no knee, as seen in **Figure 5**, but is present in the other videos. Knee heights are not evenly distributed from symmetric knee position because equal changes above and below the knee did not yield a stable PDW.

3.4.5. Category 3: Spatial and temporal asymmetry

In this video set, our intent is to examine if spatial and temporal asymmetries such as caused by limping, partial leg paralysis (hemiplegia), or a leg prosthesis will be noticeable, that is, viewed as abnormal or uncanny. In two videos, step length is held symmetric while swing time asymmetry is created (LsTa), in two videos swing time is held symmetric while step length asymmetry is created (LaTs), and in another two videos equal amounts of step length and swing time asymmetries were created (LaTa).

3.4.6. Category 4: ROS asymmetry

Walking impairment and some prosthetics can cause asymmetries in foot roll-over shape (ROS). We included three different walking patterns with asymmetric ROS foot curves. At no ROS asymmetry, both ROS are 0.333 m (1.09 feet) in radius, whereas at 100% ROS asymmetry the left ROS is 0.333 m (1.09 feet) while the right ROS is 0.111 m (0.36 feet).

3.4.7. Category 5: Knee damping with asymmetric shank mass

Four videos are included that model damping in the right knee, which simulates a stroke gait. To compensate for the damping, four different PDW shank masses were tested. The intent was to examine if a damped (i.e., impaired, injured, damaged) knee is recognizable or abnormal. If asymmetry with a damped knee is recognizable, is it possible to remove the uncanny effect by altering the impaired gait? We attempt to alter the damped gait by imposing a shank mass asymmetry. The kinematic effects on spatial and temporal gait asymmetry can be viewed in **Figure 6**. Four videos were recorded at 0, 40, 100, and 118% shank mass asymmetry. The knee damping was chosen to be 0.275 Newton-radians, which was the highest knee damping value that allowed a stable gait pattern in the PDW.

3.5. EXPERIMENTAL SETUP AND PROTOCOL

Data collection was completed using a custom internet website. This straightforward website presents one gait video (Section 3.4) at a time, while users are able to rate the shown videos on the gait's impairment and uncanniness. The website cycles through all 25 gait videos that are shown in **Table 1** in a random order. The 25 videos consist of the recorded and PDW modeled normal walking pattern, and the 23 videos of altered gait patterns using the PDW described in **Table 1**.

While watching each walking video, participants answered two questions which were presented on the screen simultaneously, each on a 7-point Likert scale (Likert, 1932) for that video. The first question asked the participants to discretely rate the video on the impairment of the gait, asking "*How normal and unimpaired does this gait appear?*". That is, participants were asked to judge the presented videos with seven options ranging from "*Normal*" to "*Very abnormal or impaired,*" with "*A little abnormal or impaired*" at the halfway point. Similarly, the second question asked the participants to rate the shown video on the uncanniness of the gait, reading "*How eerie or uncanny does this gait appear?*". The participants were given as much time as they wanted to evaluate each video which cycled from beginning to end indefinitely. The duration of all the videos was roughly thirty seconds long, however slightly varied in length depending on gait speed.

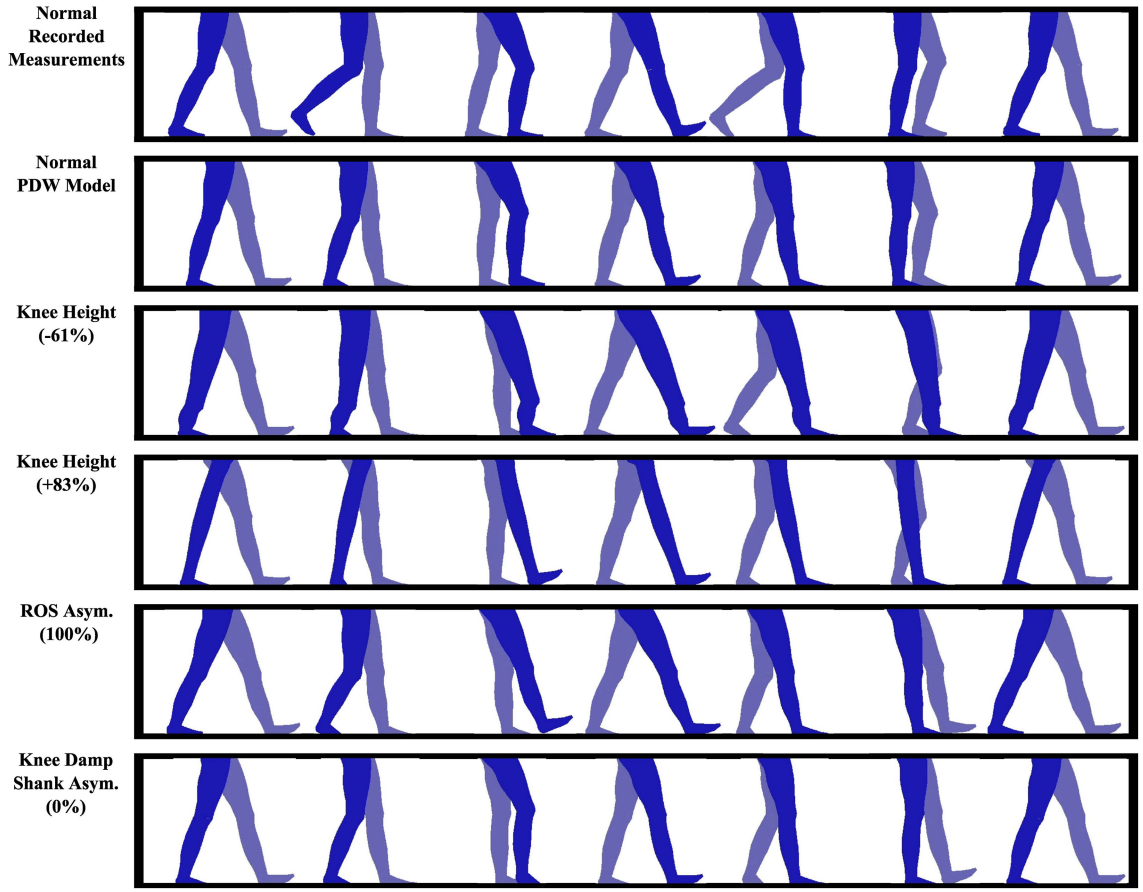


FIGURE 5 | Some of the passive dynamic walker models that were presented to participants. All videos are included in the Supplementary Material.

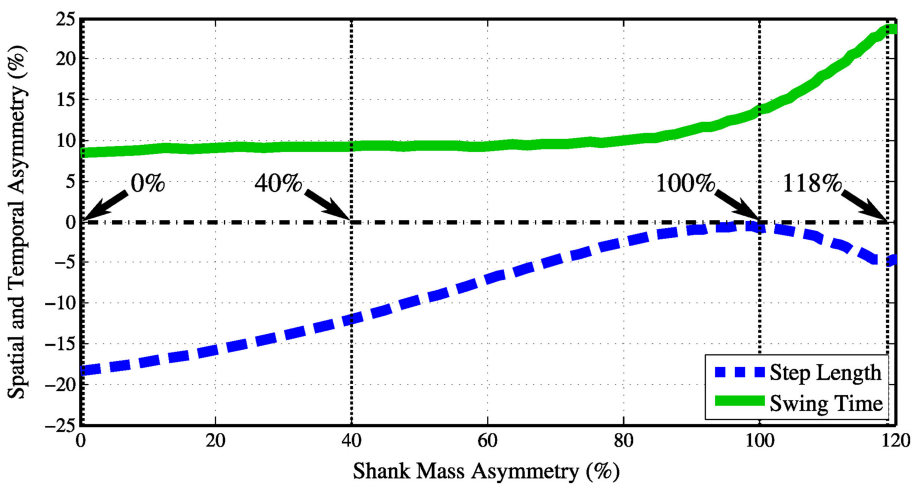


FIGURE 6 | As the right knee was damped with 0.275 Newton-radians the shank asymmetry was increased from 0 to 120%. The step length and swing time asymmetries are on opposite legs.

The participants are asked to perform at least 25 video ratings, however some participants completed as many as 144 video ratings with a median of 70 video ratings. A total of 1582 video ratings were submitted by 42 participants, however, to improve consistency, only participants that rated 4 or more gait videos were considered, which yielded 33 valid participants. Furthermore, if a video was rated twice by a participant, only the first perception score was included, reducing a potential bias from individuals that rated a video multiple times. Each video was rated a minimum of 26 times with a median of 61 video ratings per gait video. All of the videos shown to the participants are included as Supplementary Material.

The web page includes simple instructions and a clear link to an approved minimal risk University of South Florida Institutional Review Board (IRB) consent form with a waiver of documentation of consent.

3.6. STATISTICAL ANALYSIS AND EVALUATION

Participants rated walking videos on a symmetric 7-point Likert scale. Because independent participants evaluate the walking videos and the ranked quantitative responses hold true throughout the Likert scale range, we assume a continuous linearity between Likert scale points and treat the acquired data as ordinal interval-level. A Chi-square goodness-of-fit test revealed that the comprehensive data does not follow a normal distribution [$\chi^2_{(6,\,N=33)} = 844$, $p < 0.001$]. Data within each category was also found not to follow a normal distribution, where the statistics of each video category Chi-squared will be included in the following results section. Because the data for each category of videos does not follow a normal distribution, we use a Kruskal-Wallis one-way analysis of variance non-parametric test to verify whether video ratings within each category of videos originated from the same distribution (i.e., are they statistically significantly the same). A Kruskal-Wallis one-way analysis of variance by ranks test (a.k.a. Kruskal-Wallis H test or Dunn's test) is a rank-based nonparametric multiple comparison test. This *post-hoc* test is used to determine if there are statistically significant differences between two or more videos in each video category rated on the 7-point Likert scale.

The base/control gait perception rating for this experiment is the measured normal walking pattern. Before we are able to asses judgment on the impairment and uncanniness of all the videos, we first set out to compare our PDW modeled normal gait to the measured gait and determine if participants viewed our modeled walking pattern as being as normal as the recorded gait. To evaluate the statistical significance between the recorded and normal walking video, we apply a Wilcoxon rank-sum test used for nonparametric testing of the null hypothesis that the two compared populations stem from the same population. This test will be used to evaluate how close to the actual recorded normal gait the PDW modeled normal gait is.

4. RESULTS

4.1. PERCEPTION OF IMPAIRMENT

In this section, The perception of all 25 gaits in terms of gait impairments was analyzed, that is, participants' perception of the gaits' pathological nature. The results of each category are shown in **Figure 7**. All videos in each category were compared to normal gait, that is, comparison statistics included PDW modeled gait perception results for each category.

4.1.1. Normal gait

A Wilcoxon rank-sum test was used to evaluate the difference in the responses of our 7-Likert scale question on gait impairment. We found a non-significant effect between the two data sets, thus no statistically significant difference was found. The mean ranks of the recorded and modeled gait data sets were 142 and 162, respectively; $Z = 1.12$, $p > 0.05$. The number of collected ratings for the recorded and PDW modeled gait pattern was $n_1 = 26$ and $n_2 = 294$, respectively. The medians of the recorded and modeled data were both 6 as seen in **Figure 7**.

4.1.2. Category 1: Gait cadence

Chi-squared goodness of fit analysis for this category revealed that the data did not follow a normal distribution [$\chi^2_{(6,\,N=33)} = 400$, $p < 0.001$]. Nonparametric Kruskal-Wallis one-way analysis of Variances showed statistically significant difference between the perceived abnormality due to impairment of different gait cadence [$H_{(4,\,466)} = 24.4$, $p < 0.001$]. *Post hoc* analysis showed that participants were able to spot that there was something abnormal and altered between the modeled normal walking pattern and a gait that is 50 and 50% faster. However, participants were not able to statistically significantly distinguish the normal from gaits slowed down 25% and sped up 25% within this category.

4.1.3. Category 2: Knee height

Data sets in this category were found to not follow a normal distribution [$\chi^2_{(6,\,N=33)} = 317$, $p < 0.001$]. A statistically significant difference was detected [$H_{(6,\,686)} = 327.6$, $p < 0.001$]. Participants perceived all presented knee location changes as statistically significantly different compared to the normal gait. Participants evaluated knee heights of +83, +57, −40, and −61% as noticeably and highly abnormal or impaired, measuring their median, averages, and confidence intervals below neutral (4). Knee heights of +22 and −26% were only perceived as moderately impaired, which indicates that some knee height asymmetry with spatial and temporal gait symmetry could be dismissed as somewhat normal by observers. Participants were slightly more consistent in rating a low knee height as abnormal compared to higher knee locations (based on the confidence interval range).

4.1.4. Category 3: Spatial and temporal asymmetry

Data sets in this category were found to not follow a normal distribution [$\chi^2_{(6,\,N=33)} = 397$, $p < 0.001$]. A statistically significant difference was found within this category group [$H_{(6,\,689)} = 146$, $p < 0.001$]. *Post-hoc* analysis revealed both step length (L) and swing time (T) left-right asymmetries produced statistically significant differences compared to normal gait when a 13% asymmetry was imposed, however at 5% asymmetry the gait was not perceived as impaired. That is, participants did not see small independent changes in swing time and step length as impaired. The gait was perceived as recognizably impaired at 13% step length asymmetry (LaTs) (mean rank = 238), while being perceived as yet more impaired at 13% swing time asymmetry (LsTa)

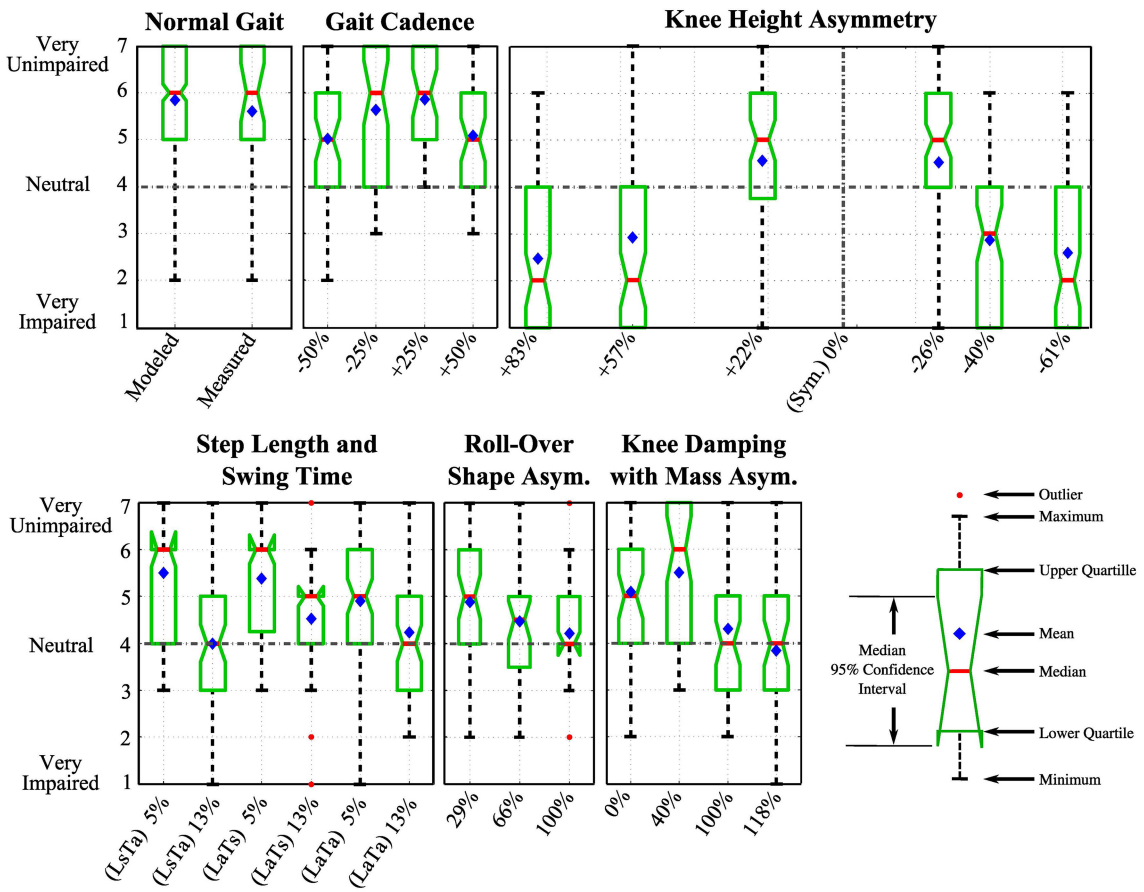


FIGURE 7 | Extended box and whiskers notch plots show participant's responses to videos in each category in response to the question: *"How normal and unimpaired does this gait appear?"*.

(mean rank = 193). However, the difference in impairment perception between these two videos was not statistically significantly different (Wilcoxon $Z = 0.55$, $p = 0.58$).

4.1.5. Category 4: ROS asymmetry

The data sets in this category did not follow a normal distribution [$\chi^2_{(6,\,N=33)} = 342$, $p < 0.001$], while a statistically significant difference among videos in this category was found [$H_{(3,\,401)} = 68$, $p < 0.001$]. Post hoc analysis showed participants perceived all videos in this category with a statistically significant difference compared to the normal gait. Walking videos with 29, 66, and 100% ROS asymmetry were perceived as minimally (mean rank = 160), moderately (mean rank = 127), and highly impaired (mean rank = 102), respectively.

4.1.6. Category 5: Knee damping with asymmetric shank mass

The data sets in this category did not follow a normal distribution [$\chi^2_{(6,\,N=33)} = 262$, $p < 0.001$], while a statistically significant difference among videos in this category was found [$H_{(4,\,462)} = 89$, $p < 0.001$]. Participants perceived all but one (40%) shank asymmetry with knee damping videos in this category with a statistically significant difference compared to the normal gait, seen in **Figure 6**. Although the 40 and 100% shank mass asymmetry had similar temporal asymmetries, 9.2 and 13%, receptively, only the 100% shank asymmetry was perceived as significantly different from normal gait. However, this may be caused by the spatial asymmetry in gait, which was 12 and 0% for the two videos respectively. Once the temporal asymmetry increased to 24% with a 4.5% spatial asymmetry, the perception of impairment was at its maximum.

4.2. PERCEPTION OF UNCANNINESS

Here the results of participants' perception of 25 gaits in terms of how uncanny (eerie or strange) the gaits appear are presented. The results for this second metric for each category are shown in **Figure 8**.

4.2.1. Normal gait

As with the impaired perception metric, we initially compare a measured and PDW modeled normal walking pattern. Wilcoxon rank-sum test yielded a non-significant effect between the two

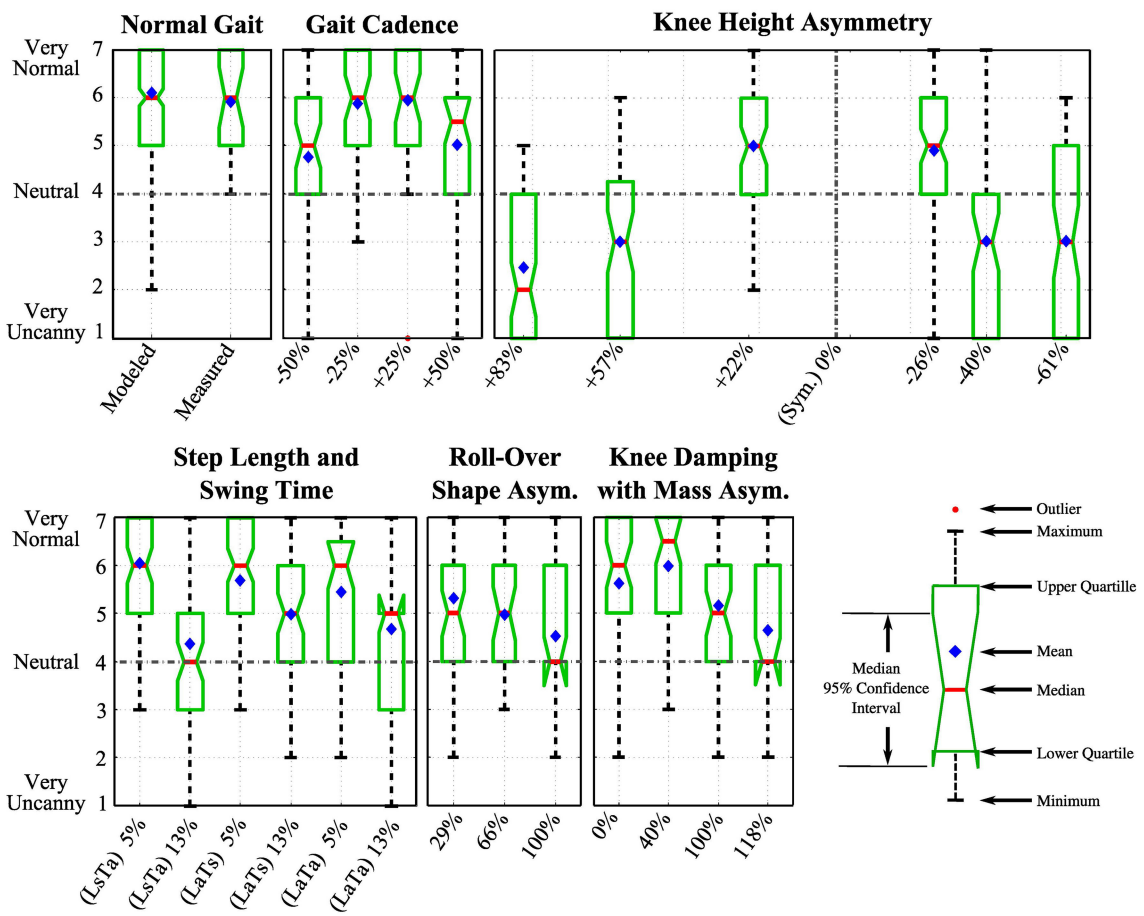


FIGURE 8 | Extended box and whiskers notch plots show participant's responses to videos in each category in response to the question: *"How eerie or uncanny does this gait appear?"*.

data sets, thus no statistically significant difference was found. The mean ranks of the recorded and modeled gait data sets were 142 and 162, respectively; $Z = 1.10, p > 0.05$. The number of collected ratings for the recorded and PDW modeled gait pattern was $n_1 = 26$ and $n_2 = 294$, respectively. The perception rating medians of the uncanniness of the recorded gait was 6, while the median rating for the modeled normal video was also 6.

4.2.2. Category 1: Gait cadence

The participant rating data did not follow a normal distribution [$\chi^2_{(6,\,N=33)} = 214, p < 0.001$]. We found a statistically significant difference between the perceived abnormality due to impairment of different gait cadence [$H_{(4,\,466)} = 47, p < 0.001$]. Participants found something more uncanny about the gait that was 50% slower and 50% faster, however median Likert score for both these altered gaits was 5 and 5.5, respectively. This shows that participants saw a slight uncanniness compared to the normal modeled walking pattern, but could not definitely say that it was uncanny. Participants were not able to statistically significantly distinguish the normal from gaits slowed down 25% and sped up 25% within this category.

4.2.3. Category 2: Knee height

Data sets in this category were found to not follow a normal distribution [$\chi^2_{(6,\,N=33)} = 317, p < 0.001$]. Among them a statistically significant difference was detected [$H_{(6,\,686)} = 327.6, p < 0.001$]. Participants perceived all presented knee location changes as statistically significantly different compared to the normal gait. In other words, individuals rated all deviations from normal gait as uncanny to some degree.

Participants evaluated knee heights of +83, +57, −40, and −61% as noticeably uncanny, measuring their median, averages, and confidence intervals below the neutral score of 4, however knee heights of +22 and −26% were only perceived as moderately impaired and below a neutral uncanny perception. These results are very similar to the previously discussed impairment ratings with the exception of knee height asymmetry of +57 and −40% which was rated a 3 instead of a 2.

4.2.4. Category 3: Spatial and temporal asymmetry

Data sets in this category were found to not follow a normal distribution [$\chi^2_{(6,\,N=33)} = 329, p < 0.001$]. A statistically significant difference was found within this category group [$H_{(6,\,689)} =$

135, $p < 0.001$]. As opposed to the impairment ratings, uncanny ratings were consistently lower, however the same trends persisted. Both step length (L) and swing time (T) left-right asymmetries produced statistically significant differences in uncanniness compared to normal gait when a 13% asymmetry was imposed, however at 5% asymmetry the gait was not perceived as impaired. That is, participants did not see small independent changes in swing time and step length as uncanny, while larger temporal and spatial asymmetries each at 13% were perceived as recognizably more more uncanny at medium ranks at 4 and 5, respectively.

4.2.5. Category 4: ROS asymmetry

With all the previous groups, the collected ratings for this category was a normal distribution [$\chi^2_{(6,\,N=33)} = 321$, $p < 0.001$], while a statistically significant difference among videos in this category was found [$H_{(3,\,401)} = 66$, $p < 0.001$]. As with the gait impairment perception, post hoc analysis showed participants perceived all videos in this category with a statistically significant difference compared to the normal gait. Walking videos with 29, 66, and 100% ROS asymmetry were perceived as minimally (mean rank = 163), moderately (mean rank = 132), and highly impaired (mean rank = 100), respectively.

4.2.6. Category 5: Knee damping with asymmetric shank mass

Again, the data sets in this category did not follow a normal distribution [$\chi^2_{(6,\,N=33)} = 375$, $p < 0.001$], while a statistically significant difference among videos in this category was found [$H_{(4,\,462)} = 55$, $p < 0.001$]. Participants perceived all but two (0 and 40%) shank asymmetry with knee damping videos in this category with a statistically significant difference compared to the normal gait. As seen in **Figure 6**, the same trends arise as in the impairment rating results, with the exception that 0% shank mass asymmetry with knee damping was not significantly different then a normal walking pattern.

5. DISCUSSION

Statistically, participants were shown not to be able to effectively differentiate between a recorded healthy human gait and a modeled PDW walking pattern in terms of impairment and uncanniness. Hence, it was viable to compare a modeled PDW gait to further walking models that have been systematically altered. This also suggests that there are significant visual characteristics of PDW gaits that are similar to human gaits as is expected since the kinematics are similar (Donelan et al., 2002; Handžić and Reed, 2013b). Although the trend was similar to impairment ratings, uncanny rating confidence intervals were generally shifted slightly toward normal perception.

Trends of the perception on gait impairment is similar to the perception on gait uncanniness, however the uncanny perception seems generally slightly and consistently closer to normal PDW gait than the impairment perception of the same walking pattern. This means that the participants consistently recognized abnormal gaits as pathological, but did not feel an equally strong uncanny or eerie feeling while watching the gait. Thus, we believe most of the perceptions in this study are along the top-right portion of the human-likeness/familiarity shown in **Figure 1**.

The combined results of this study confirms the conclusions drawn by Thompson et al. (2011) and Piwek et al. (2014), which strengthen the counterclaims against an uncanny valley for computer generated synthetic human body motions. Similar to their results, we were only able to find monotonically decreasing familiarity with heightened abnormality. However, we can only speculate about why. Our abnormality (stimuli) resolution could have been too low to find a valley dip. In addition, the amplitude of the imposed abnormality may not have been substantial enough to map it onto the most right side of the uncanny valley. Furthermore, our study only focused on the lower extremity kinematics, hence in the light of this focus and previous studies, the effects of the uncanny valley may be minimal or even nonexistent.

Normal gait with a gait cadence increased or decreased by 50% was noticed as slightly more impaired and uncanny when compared to a normal gait cadence. This may indicate that when seeing someone walking hastily or abnormally slow, it can be interpreted as out of the ordinary and draws attention, signaling that some impairment or abnormalities are present. Although both the impairment and uncanniness of these videos were significantly different than the normal walking pattern, participants' medium rating hovered between 5 and 6, that is, neutral to very unimpaired/uncanny. Such a reaction may draw some attention from observers, however would generally not be considered abnormal.

A inverse "V" pattern shows the increase of participants' gait impairment and uncanniness perception with knee height asymmetry, with a focal area between +10 and −20% knee height change. These results imply that given step length and step time symmetry, some knee height asymmetry can be unrecognizable or even perceived as normal. As opposed to the other categories, alteration of knee height symmetry provoked the highest participant impairment or uncanny ratings. It is shown that the higher the knee location is moved from its symmetric position, the more the gait is perceived as impaired or even uncanny. These results also suggest that a prosthetic design with a lowered knee location for functional improvement (Sushko et al., 2012; Ramakrishnan, 2014) may be unnoticeable to some extent. It should be noted that the experiment did not examine if or how clothing and wearing loose-fitting clothes would help to hide the effect of a prosthetic with a knee location in a different location, but these effects are likely to mask the knee location.

Separately, 5 and 5% LaTs did not produce a perception of impairment or uncanniness with participants, inherently suggesting that some gait asymmetry is not noticeable by observers and it should be noted that healthy individuals are known to have some asymmetric gait parameters (Sadeghi et al., 2000). For example, for observers to consistently not notice gait asymmetry such as a limb caused by a prosthetic or injury, one can walk with a 5% spatial or temporal asymmetry. However, it is interesting to note that 5% simultaneously in both measures produces a moderate perception of abnormality but with the confidence interval below the neutral perception rating. It may be concluded that compounding these asymmetries may cause greater perceptions in abnormality, however this seems not to be the case for 13% LaTa. The 13% LaTa was rated similar to the 13% LaTs, while 13%

LsTa was rated more impaired and uncanny than 13% LaTa. A further study using more combinations of these asymmetric gait measures would help to understand the perceptual interactions with gait asymmetry more fully.

Although more ROS asymmetries would clarify a trend, it can be concluded that with all factors symmetric, a ROS asymmetry below around 35% can pass as minimally impaired or uncanny by observers. The trend implies that a ROS asymmetry below 15% may not be distinguishable from a normal and healthy gait. This is not surprising since ROS have enormous effects on gait kinematics, kinetics, and balance (Menant et al., 2009). ROS are important in prosthetic design (Hansen et al., 2000; Curtze et al., 2009) and for reducing forces on the user's stump (Rietman et al., 2002). Specially-designed shoe soles can also benefit individuals with cerebral palsy, Parkinsons, and stroke (Rodriguez and Aruin, 2002).

Category 5 results imply that if a person suffering from an impairment causing damping in a knee (injury, neurological, etc.), that person could be seen as impaired or even slightly uncanny. However, imposing an accompanying asymmetry, such as adding an asymmetric mass distribution, can potentially alleviate the perception of impairment or uncanniness. In other words, as one gait asymmetry is imposed that causes gait perception of impairment and uncanniness, a second gait asymmetry may be applied to some degree to negate these perceptions. This combination of asymmetries could lead to gait patterns that balance the perceptual and dynamic aspects of gait.

Results from Category 5 agree with the conclusions drawn from Category 3 since, looking at **Figure 6**, it can be concluded that the swing time asymmetry has a greater effect on participants noticing the abnormality than step length asymmetry. Although a person may step with symmetric step distances, the difference in limb swing time is far more noticeable to observers as shown with gait of Category 5 100% shank mass asymmetry and Category 3 13% LsTa. It is interesting to note that these results are comparable to Lauzière et al. (2014) who looked into the perception of ones own gait asymmetry (internal), which concluded that the parameter that corresponded the most with belt speed asymmetry was found to be stance time.

This normalizing of the perception of joint damping can also potentially be achieved by altering other gait parameters such as having a foot roll-over shape or knee height asymmetry, however, this is still open for future studies.

6. CONCLUSIONS AND FUTURE WORK

In this study we outlined the boundaries of perceived gait impairment and uncanniness of some pathological or altered gait patterns including moved knee height and asymmetric foot roll-over shape radii. Despite a selected number of gait alteration parameters, we were able to explore the perception of pathological or uncanny gait. Generally, perception rating trends were the same between impaired and uncanny ratings, however the uncanny rating was consistently more normal. This similarity in trends may suggest a coupling between the perception of impairment and the uncanny. Although we have shown that altering human gait parameters alters the perception of normal and healthy walking to observers, further investigation with different types of gait pathologies and a greater resolution of abnormalities in walking patterns for each category is needed.

We conclude that there clearly is a gray and undefined area in human perception in gait, where human gait may be abnormal while being perceived as unimpaired or uncanny. The gait abnormalities that we analyzed were gait cadence, knee height asymmetry, spatial and temporal walking asymmetries, and foot roll-over shape asymmetry. We also examined the perception of gait by changing two independent gait parameters, specifically asymmetric knee stiffness and shank mass asymmetry. This multi-parameter analysis clearly showed that it is possible to alter the perception of a gait impairment by manipulating different gait parameters. These results are promising and such a multi-parameter manipulation technique may be useful in the field of prosthetic or hemiplegic gait analysis and rehabilitation, in that a noticeable gait asymmetry could be hidden by imposing and altering other gait parameters. Although promising, further investigation of a more clear relationship between manipulating multiple gait parameters and the effect on gait perception is still to be researched.

Future work on this study includes a larger scale public video rating system such as the one presented in this study, however with more videos covering a larger range of parameters such as further variation of knee location with a finer abnormality amplitude resolution. Although, we have moved one knee location to an asymmetric position, it would be interesting to examine if moving both knees equal distances provoke the same reactions in participants. These videos may also include studying the effects of altered limb thickness, texture, or limb form. We believe that the results of this and further investigations of what is considered normal human gait can help researchers, designers, and developers of gait modification devices, such as prosthetics or joint braces, create functionally better and more socially accepted devices. In this study we were only considering deviation of walking cadence and various parameter asymmetries, however further quantitative and qualitative investigation in the perception of the way the limbs move, that is, the limb angle trajectories (position, velocity, etc.).

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. MRI-1229561.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpsyg.2015.00199/abstract

A Supplementary video of all gait patterns presented to participants is provided along with this manuscript. All the collected data and a Matlab® programming code used for interpretation is included as well. A snapshot image of the internet web page used to collect all perception data is included as well.

REFERENCES

Adamczyk, P. G., Collins, S. H., and Kuo, A. D. (2006). The advantages of a rolling foot in human walking. *J. Exp. Biol.* 209, 3953–3963. doi: 10.1242/jeb.02455

Anderson, M., Jenkins, O., and Osentoski, S. (2011). Recasting robotics challenges as experiments [competitions]. *Robo. Autom. Mag. IEEE* 18, 10–11. doi: 10.1109/MRA.2011.941627

Blake, R., and Shiffrar, M. (2007). Perception of human motion. *Annu. Rev. Psychol.* 58, 47–73. doi: 10.1146/annurev.psych.57.102904.190152

Bohannon, R. W., Andrews, A. W., and Smith, M. B. (1988). Rehabilitation goals of patients with hemiplegia. *Int. J. Rehabil. Res.* 11, 181–184. doi: 10.1097/00004356-198806000-00012

Bohannon, R. W., Morton, M. G., and wikholm, J. B. (1991). Importance of four variables of walking to patients with stroke. *Int. J. Rehabil. Res.* 14, 246–250. doi: 10.1097/00004356-199109000-00010

Chen, V. F. H. (2005). *Passive Dynamic Walking with Knees: A Point Foot Model*. Master's thesis, Massachusetts Institute of Technology.

Curtze, C., Hof, A. L., van Keeken, H. G., Halbertsma, J. P., Postema, K., and Otten, B. (2009). Comparative roll-over analysis of prosthetic feet. *J. Biomech.* 42, 1746–1753. doi: 10.1016/j.jbiomech.2009.04.009

Donelan, J. M., Kram, R., and Kuo, A. D. (2002). Mechanical work for step-to-step transitions is a major determinant of the metabolic cost of human walking. *J. Exp. Biol.* 205, 3717–3727. Available online at: http://jeb.biologists.org/content/205/23/3717.abstract

Drillis, R., Contini, R., and Bluestein, M. (1964). *Body Segment Parameters: A Survey of Measurement Techniques*. Vol. 8 (Washington, DC: National Academy of Sciences), 44–66.

Eberle, S. G. (2012). "Play: a polyphony of research theories, and issues," in *Pinpointing Play at the Edge of the Uncanny Valley,* Vol. 12 (Lanham, MD: University Press of America), 133–162.

Freud, S. (1919). *The Uncanny (J. Strachey, Trans.). The Standard Edition of Complete Psychological Works of Sigmund Freud*, Vol. 17 (London: The Hogarth Press).

Giese, M., and Lappe, M. (2002). Measurement of generalization fields for the recognition of biological motion. *Vis. Res.* 42, 1847–1858. doi: 10.1016/S0042-6989(02)00093-7

Grabowski, A. M., McGowan, C. P., McDermott, W., Beale, M., Kram, R., and Herr, H. (2010). Running-specific prostheses limit ground-force during sprinting. *Biol. Lett.* 6, 201–204. doi: 10.1098/rsbl.2009.0729

Handžić, I., and Reed, K. B. (2013b). Validation of a passive dynamic walker model for human gait analysis. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 6945–6948. doi: 10.1109/EMBC.2013.6611155

Handžić, I., and Reed, K. B. (2013a). Comparison of the passive dynamics of walking on ground, tied-belt and split-belt treadmills, and via the gait enhancing mobile shoe (GEMS). *IEEE Int. Conf. Rehabil. Robot.* doi: 10.1109/ICORR.2013.6650509

Hansen, A., Childress, D., and Knox, E. (2000). Prosthetic foot roll-over shapes with implications for alignment of trans-tibial prostheses. *Prosthet. Orthot. Int.* 24, 205–215. doi: 10.1080/03093640008726549

Hansen, A., and Wang, C. (2010). Effective rocker shapes used by able-bodied persons for walking and fore-aft swaying: Implications for design of ankle-foot prostheses. *Gait Posture* 32, 181–184. doi: 10.1016/j.gaitpost.2010.04.014

Hansen, A. H., and Childress, D. S. (2004). Effects of shoe heel height on biologic rollover characteristics during walking. *J. Rehabil. Res. Dev.* 41, 547–554. doi: 10.1682/JRRD.2003.06.0098

Henderson, G., and Bryan, W. V. (2004). *Psychosocial Aspects of Disability*. Springfield, IL: Charles C Thomas Publisher.

Honeycutt, C., Sushko, J., and Reed, K. B. (2011). Asymmetric passive dynamic walker. *IEEE Int. Conf. Rehabil. Robot.* 852–857. doi: 10.1109/ICORR.2011.5975465

Jentsch, E. (1906). Zur psychologie des unheimliche. *Psychiatrisch-Neuroligische Wochenshrift* 8, 195–198.

Knoblich, G., and Flach, R. (2001). Predicting the effects of actions: Interactions of perception and action. *Psychol. Sci.* 12, 467–472. doi: 10.1111/1467-9280.00387

Kozlowski, L. T., and Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Percept. Psychophys.* 21, 575–580. doi: 10.3758/BF03198740

Kuo, A. D. (2007). The six determinants of gait and the inverted pendulum analogy: a dynamic walking perspective. *Hum. Mov. Sci.* 26, 617–656. doi: 10.1016/j.humov.2007.04.003

Lauzière, S., Miéville, C., Duclos, C., Aissaoui, R., and Nadeau, S. (2014). Perception threshold of locomotor symmetry while walking on a split-belt treadmill in healthy elderly individuals 1, 2, 3. *Percept. Mot. Skills* 118, 475–490. doi: 10.2466/25.15.PMS.118k17w6

Lee, J., Chai, J., Reitsma, P. S., Hodgins, J. K., and Pollard, N. S. (2002). "Interactive control of avatars animated with human motion data," in *ACM Transactions on Graphics (TOG),* Vol. 21 (New York, NY: ACM), 491–500.

Likert, R. (1932). A technique for the measurement of attitudes. *Arch. Psychol*, Vol. 22, p. 55.

Lipson, J. G., and Rogers, J. G. (2000). Cultural aspects of disability. *J. Transcult. Nurs.* 11, 212–219. doi: 10.1177/104365960001100308

Loula, F., Prasad, S., Harber, K., and Shiffrar, M. (2005). Recognizing people from their movement. *J. Exp. Psychol. Hum. Percept. Perform.* 31:210. doi: 10.1037/0096-1523.31.1.210

MacDorman, K. F., Green, R. D., Ho, C.-C., and Koch, C. T. (2009). Too real for comfort? uncanny responses to computer generated faces. *Comput. Hum. Behav.* 25, 695–710. doi: 10.1016/j.chb.2008.12.026

McDonnell, R., Breidt, M., and Bülthoff, H. H. (2012). Render me real?: investigating the effect of render style on the perception of animated virtual humans. *ACM Trans. Graph.* 31, 91. doi: 10.1145/2185520.2185587

McDowell, M., Fryar, C., Ogden, C., and Flegal, K. (2008). Anthropometric reference data for children and adults: United States, 2003-2006. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, 2008.

McGeer, T. (1990). Passive dynamic walking. *Int. J. Robo. Res.* 9, 62–82. doi: 10.1177/027836499000900206

Menant, J. C., Steele, J. R., Menz, H. B., Munro, B. J., and Lord, S. R. (2009). Effects of walking surfaces and footwear on temporo-spatial gait parameters in young and older people. *Gait Posture* 29, 392–397. doi: 10.1016/j.gaitpost.2008.10.057

Mori, M. (1970). The uncanny valley. *Energy* 7, 33–35.

Perry, J. (2010). *Gait Analysis: Normal and Pathological Function, 2nd Edn.*, Vol. 50. Thorofare, NJ: SLACK Inc.

Piwek, L., McKay, L., and Pollick, F. (2014). Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition* 130, 271–277. doi: 10.1016/j.cognition.2013.11.001

Rabufetti, M., Recalcati, M., and Ferrarin (2005). Trans-femoral amputee gait: Socket-pelvis constraints and compensation strategies. *Prosthet. Orthot. Int.* 29, 183–192. doi: 10.1080/03093640500217182

Raibert, M. (2010). "Dynamic legged robots for rough terrain," in *Humanoid Robots (Humanoids), 2010 10th IEEE-RAS International Conference* (Nashville, TN). doi: 10.1109/ICHR.2010.5686280

Ramakrishnan, T. (2014). *Asymmetric Unilateral Transfemoral Prosthetic Simulator.* Master's thesis, University of South Florida.

Rietman, J., Postema, K., and Geertzen, J. (2002). Gait analysis in prosthetics: opinions, ideas and conclusions. *Prosthet. Orthot. Int.* 61, 50–57. doi: 10.1080/03093640208726621

Rodriguez, G., and Aruin, A. (2002). The effect of shoe wedges and lifts on symmetry of stance and weight bearing in hemiparetic individuals. *Arch. Phys. Med. Rehabil.* 83, 478–482. doi: 10.1053/apmr.2002.31197

Sadeghi, H., Allard, P., Prince, P., and Labelle, H. (2000). Symmetry and limb dominance in able-bodied gait: a review. *Gait Posture* 12, 34–45. doi: 10.1016/S0966-6362(00)00070-9

Sakagami, Y., Watanabe, R., Aoyama, C., Matsunaga, S., Higaki, N., and Fujimura, K. (2002). "The intilligent ASIMO: system overover and integration," in *International Conference on Intelligent Robots and Systems* (Lausanne: IEEE), 2478–2483.

Schmalz, T., Blumentritt, S., and Jarasch, R. (2002). Energy expenditure and biomechanical characteristics of lower limb amputee gait: the influence of prosthetic alignment and different prosthetic components. *Gait Posture* 16, 255–263. doi: 10.1016/S0966-6362(02)00008-5

Seyama, J., and Nagayama, R. S. (2007). The uncanny valley: effect of realism on the impression of artificial human faces. *Presence* 16, 337–351. doi: 10.1162/pres.16.4.337

Sushko, J., Honeycutt, C., and Reed, K. B. (June 2012). "Prosthesis design based on an asymmetric passive dynamic walker," in *Biomedical Robotics and Biomechatronics (BioRob), 2012 4th IEEE RAS and EMBS International Conference* (Rome), 1116–1121.

Thompson, J., Trafton, G., and McKnight, P. (2011). The perception of humanness from the movements of synthetic agents. *Perception* 40, 695–704. doi: 10.1068/p6900

Whittle, M. (2012). *Gait Analysis, 5th Edn*. London: Elsevier Health Sciences.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 April 2014; accepted: 09 February 2015; published online: 27 February 2015.

*Citation: Handžić I and Reed KB (2015) Perception of gait patterns that deviate from normal and symmetric biped locomotion. Front. Psychol. **6**:199. doi: 10.3389/fpsyg.2015.00199*

This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 25 February 2015
doi: 10.3389/fpsyg.2015.00204

Walking in the uncanny valley: importance of the attractiveness on the acceptance of a robot as a working partner

Matthieu Destephe[1]*, Martim Brandao[2], Tatsuhiro Kishi[2], Massimiliano Zecca[3,4,5], Kenji Hashimoto[1] and Atsuo Takanishi[1,6]

[1] *Department of Modern Mechanical Engineering, Waseda University, Tokyo, Japan*
[2] *Graduate School of Advanced Science and Engineering, Waseda University, Tokyo, Japan*
[3] *School of Electronic, Electrical and Systems Engineering, Loughborough University, Loughborough, UK*
[4] *National Centre for Sports and Exercise Medicine – East Midlands, Loughborough, UK*
[5] *Leicester-Loughborough Diet, Lifestyle and Physical Activity Biomedical Research Unit, National Institute for Health Research, Loughborough, UK*
[6] *Humanoid Robotics Institute, Waseda University, Tokyo, Japan*

Edited by:
Marcus Cheetham, University of Zurich, Switzerland

Reviewed by:
Danielle Sulikowski, Charles Sturt University, Australia
Julia Fink, Ecole Polytechnique Fédérale de Lausanne, Switzerland
Shuichi Nishio, Advanced Telecommunications Research Institute International, Japan

****Correspondence:***
Matthieu Destephe, Department of Modern Mechanical Engineering, Waseda University, Kikuicho Campus 41-5-3F-03/4, 17 Kikuicho, Shinjuku, Tokyo, Japan
e-mail: matthieu@takanishi.mech.waseda.ac.jp

The Uncanny valley hypothesis, which tells us that almost-human characteristics in a robot or a device could cause uneasiness in human observers, is an important research theme in the Human Robot Interaction (HRI) field. Yet, that phenomenon is still not well-understood. Many have investigated the external design of humanoid robot faces and bodies but only a few studies have focused on the influence of robot movements on our perception and feelings of the Uncanny valley. Moreover, no research has investigated the possible relation between our uneasiness feeling and whether or not we would accept robots having a job in an office, a hospital or elsewhere. To better understand the Uncanny valley, we explore several factors which might have an influence on our perception of robots, be it related to the subjects, such as culture or attitude toward robots, or related to the robot such as emotions and emotional intensity displayed in its motion. We asked 69 subjects ($N = 69$) to rate the motions of a humanoid robot (*Perceived Humanity*, *Eeriness*, and *Attractiveness*) and state where they would rather see the robot performing a task. Our results suggest that, among the factors we chose to test, the attitude toward robots is the main influence on the perception of the robot related to the Uncanny valley. Robot occupation acceptability was affected only by *Attractiveness*, mitigating any Uncanny valley effect. We discuss the implications of these findings for the Uncanny valley and the acceptability of a robotic worker in our society.

Keywords: humanoid robot, emotion, uncanny valley, cross-cultural study, acceptability

INTRODUCTION

As Robotics as a science progresses, robots develop improved functionalities. The DARPA (Defense Advanced Research Projects Agency) Robotics Challenge is bringing highly sophisticated robots, mainly humanoids, to the disaster theaters to help humans and assist in rescues. Besides rescuers, humanoid robots may have other roles, especially in our aging society: nurse, receptionist, nanny, house helper, or even kindergarten teacher. When building robots to help or service us, it is important to understand what makes a robot acceptable. For example, the personality of the robot has to adapt to the job itself and not to the users' personality in order to have a higher social trust from them to complete a certain task or job (Joosse et al., 2013). Also, some jobs are favored for robots and some for humans (Takayama et al., 2008). Whenever memorization, acute perception and service to others are the main features of a job description, people would be comfortable to have a robot doing the job. Whenever artistic creation, evaluation, judgment, and diplomacy are required people would prefer a human performing the job.

Most of the service jobs entail a form of emotion regulation which is called Emotional labor (Hochschild, 2003). Emotion labor jobs require face-to-face interaction with customers and influence their emotional state. In face-to-face interactions, displaying emotions—and sometimes not displaying them or tuning them down—helps the outcomes of said interactions (Dasborough and Ashkanasy, 2002; Prati et al., 2003). For jobs where emotional labor is necessary, would the use of an emotional robot would be adequate (i.e., would the robot transmit the correct message and influence the person it is interacting with in an appropriate way) and more importantly, not provoke feelings of unease? When humanoid robots are designed to interact with people, there is a risk of rejection from the users due to the robots similarity with humans. A hypothesis called "the Uncanny Valley," quite popular in Human-Robot Interaction (HRI), tries to explain this phenomenon. Developed by Mori in 1970, the Uncanny valley phenomenon occurs when the more human-like a thing is (a doll, a robot, etc.,) the more familiar people feel toward that thing (Mori, 1970). Nonetheless this relationship is

not linear: when human-likeness is close to perfect but some differences still exist, the curve collapses and the feeling, which was familiar, becomes uncanny. The term uncanny is the English translation of the German *Unheimlich*, a word describing something being felt simultaneously as familiar, strange, and scary. When the human-likeness reaches the point where it is quite hard to tell the difference from a human being, the curve rises steeply again, outlining the shape of a valley, thus giving the name "Uncanny valley" to that phenomenon (**Figure 1**) (MacDorman et al., 2005).

Despite its popularity, there is still uncertainty about what would be the cause of this phenomenon. Some recent studies do not support the existence of the Uncanny valley (Bartneck et al., 2009a; Thompson et al., 2011) as they found little to no evidence of the expected results. However, other studies (Ho et al., 2008; Mitchell et al., 2011) support the existence of the phenomenon. Researchers have tried to understand the disparity of the results (Pollick, 2010) but so far no consensus has been reached.

Many studies have been done on the effect of robots' appearance and even some were done on robot movement (Pollick, 2010). Nonetheless the Uncanny valley phenomenon was not studied with humanoid robots expressing emotions with different intensities. As robot developers, we see several limitations in the few studies using robot motions to test the Uncanny valley phenomenon. The first issue is to use of the same movement with different media (human, human-like robot, and machine-like robot) such as performed by Saygin et al. (2012). They discovered that android motions increase brain activity in the action perception system compared to human or robot motions. Nonetheless they found that the repetition suppression effects were stronger for the human-like robot indicating a possible neural basis for the Uncanny valley phenomenon. While this kind of study is interesting *per-se*, it does not inform us about how to improve robot motions and make them more acceptable for users. The second issue is the use of a wide range of different robots performing motions without relation between them (MacDorman, 2006). In the study the author used as stimuli videos of 13 different robots performing diverse activities and found that the humanness of a robot is not the only factor influencing the eeriness perceived by the participants. However, using several robots quite dissimilar in shape and design might hinder the appearance effect and a single (or similar) motion with a neutral meaning should be used to avoid biasing the results. To overcome those limitations, we created several human-like gait patterns with different emotional intensities for a unique full-body human-sized robot and assessed them. Instead of using movements designed by an animator, we use gait data captured from experiments with professional actors (Destephe et al., 2013a). After analysis of the movements, we created for two emotions [Happiness and Sadness two patterns with different intensities (natural and exaggerated emotional intensities)]. We also created a non-emotional pattern to serve as control. Those patterns were assessed by showing videos of the humanoid robot to French and Japanese subjects. They assessed them through a specialized questionnaire (Ho's modified Godspeed questionnaire) (Ho and MacDorman, 2010) and rated their acceptability for different types of jobs.

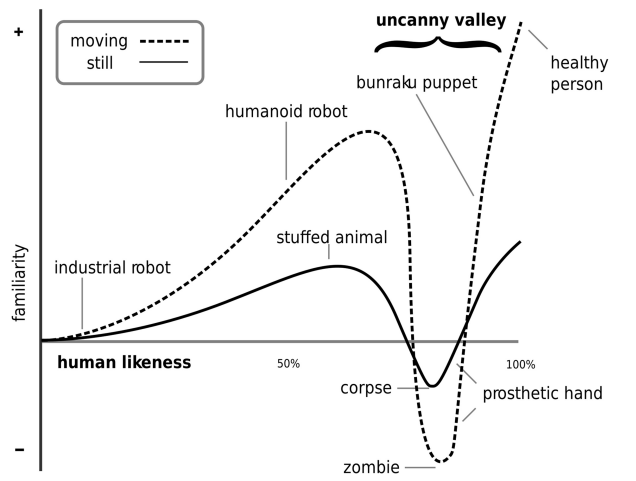


FIGURE 1 | The Uncanny valley. Mori describes the Uncanny valley as a non-linear relation between the perceived familiarity felt toward a thing and its human likeness (MacDorman et al., 2005).

We propose, in accordance with the Uncanny valley hypothesis, that as the perception of Humanness grows, the Eeriness rating follows an Uncanny valley-like shape. We predict a cultural difference in the perception of the *Eeriness* and *Attractiveness*. Japanese people and French people have a different views on what is natural or artificial (Berque, 1997; Kaplan, 2004). French people see natural things and artificial things as opposed: they see the world as hierarchical, boundaries limiting things and categories defining them. This mindset might be influenced by the Cartesian French education (Weinshall, 1971; Lubatkin et al., 2005). Inheriting a tradition of Buddhist (everything is considered to be a manifestation of same greater concept) and Shintoism (spiritual essence can be manifested in any form from rock to rivers through animals and even humans) (Earhart, 1982), Japanese people would see natural things and artificial things connected and being parts of a bigger picture. These distinctions might influence the *Eeriness* perception: Japanese participants would be less sensitive to discrepancies in the robot, thus rating lower Eeriness than French participants. Japanese people will prefer Natural Intensity emotion representation and French people will prefer Exaggerated Intensity emotion representation. The *Attitude toward robots* and the *Age* factors will predict how people perceive the robot: participants with a positive attitude and young participants (under 30 years old) will rate *Humanness*, *Attractiveness* higher, and *Eeriness* lower. Participants with a negative attitude and old participants (more than 50 years old) will rate *Humanness* and *Attractiveness* lower, and *Eeriness* higher. Finally, we hypothesize that the perception of the Uncanny valley (robot being eerie or not) will influence the participants to say whether an occupation is acceptable or not for the robot.

METHODS

PARTICIPANTS

A total of 70 subjects participated to this experiment but one was excluded from the analysis due to a software issue ($N = 69$). The participants were invited to participate to a study about

HRI through announcement in class, social network services, and mailing-lists. This study is a cross-cultural study between French and Japanese people. A total of 47 French subjects participated to this experiment (N_{FR} = 47) with an average age of 34.7 ± 12.5 years old ranging from 21 to 81 years old [28 males (33.9 y.o. ± 12.5) and 19 females (35.9 y.o. ± 12.6)]. A total of 22 Japanese subjects participated to this experiment (N_{JP} = 22) with an average age of 29.2 ± 7.1 years old ranging from 21 to 53 years old [9 males (26.3 y.o. ± 5.0) and 13 females (32.2 y.o. ± 7.9)]. The ethical committee approved the experiment protocols, the participants gave us their written informed consent and all the data collected are anonymized. The participants were recruited through on social network websites, general forums, and mailing-lists with no relation to robotics or robots.

OUR ROBOT

The videos used for our work are based on the humanoid robot WABIAN-2R (**Figure 2**). Unlike most bipedal humanoid robots, WABIAN-2R is able to perform a human-like walking with stretched knees thanks to its 2-DoF waist during the stance phase while other robots walk with bent knees (Ogura et al., 2006). WABIAN-2R is 1.5 m in height, and 64 kg in weight. Its design allows human-like gait including heel-contact and toe-off phases. This robot is mainly used for locomotion experiments and to study human movements. Besides an advanced locomotion technology, the head is a neutral, stylized human-like shape with no distinguishable features. We chose this robot because having no facial expression helps to focus on the expressivity of the whole body without having any influence coming from facial expressions.

STIMULI

The robot emotional walking patterns were created from our previous study (Destephe et al., 2013a). Two professional actors (who acted in plays, drama, and movies) were asked to perform several types of emotional walking such as Sadness, Happiness, Anger, and Fear and with different intensities: Natural (Low, Intermediate, and High) and Exaggerated. We categorize the three intensities (Low, Intermediate, and High) as Natural because the actors were asked to act in such a way that they would correspond to natural occurrences of emotion expression in daily life. The Exaggerated intensity on the other hand were performed with extravagant theatricality, broad gestures, and overplayed expressions, comparable to emotions expressions seen in plays and theaters.

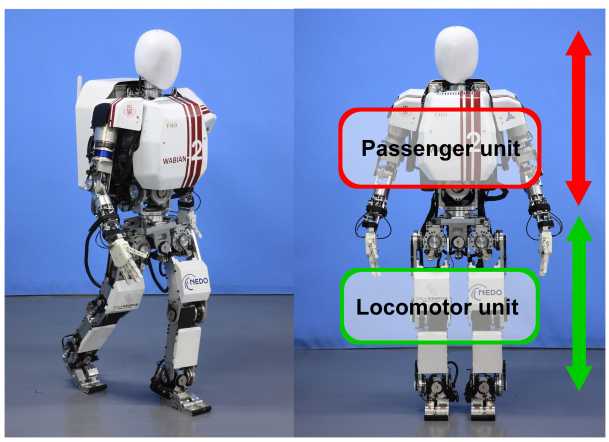


FIGURE 2 | The humanoid robot WABIAN-2R. It is capable of human-like walking with its stretched knees and pelvis mechanism.

For this work, we used Happiness and Sadness walking patterns and for each of them, one walking pattern of Natural intensity (High) and another of Exaggerated intensity. The walking patterns were based on actors' whole body movements (Destephe et al., 2013a). They were created manually such as to approximate the actors' motion as much as possible, within the constraints related to differences in the structure and the dynamics of a human body and a humanoid body. We scaled the actors' values to respect the hardware limits of the robot and used our pattern generator to generate stable walking patterns. In our previous work (Destephe et al., 2013b), the gait patterns we created achieved a high recognition rate (Natural (High) intensity/Exaggerated intensity) (Happiness: 75.0/85.7%; Sadness: 75.0/92.9%) when we assessed them in simulation with subjects. Examples of patterns used in this study are shown in the **Figure 3**.

QUESTIONNAIRE

The questionnaire we gave to the participants of this study is composed of four sub-questionnaires. For a description of the complete questionnaire, please refer to the reproduction given in the Supplementary Material.

Sub-questionnaire #1. The first sub-questionnaire asks for general information: sex, age, nationality, education level and current occupation.

Sub-questionnaire #2. The second sub-questionnaire enquires about the participant's robot-related experiences and their attitude toward robots based on the MacDorman questionnaire (MacDorman et al., 2008).

Sub-questionnaire #3. The third sub-questionnaire is described as a personality questionnaire. In fact, that questionnaire is a short screening questionnaire for autism called AQ10 (Autism spectrum Quotient with 10 items) (Allison et al., 2012).

Sub-questionnaire #4. The last sub-questionnaire assesses the participant's reactions and feelings about our emotional robot and is based on Ho's questionnaire (Ho and MacDorman, 2010). This questionnaire is designed to assess if there is the Uncanny valley phenomenon. There are several popular questionnaires used to study the reaction of robots' users such as the Godpseed questionnaire (Bartneck et al., 2009b). The Godspeed questionnaire is not well-adapted to measure the reactions to humanoid robots as several scales are redundant and it does not evaluate well the Uncanny valley phenomenon. Two issues are occurring with the Godspeed questionnaire. First, some of the semantic items do not encode well enough the indices they are related to. Second, Anthropomorphism, Likeability, Animacy, and Perceived Intelligence are highly correlated between themselves, therefore they encode the same concept (most probably the humanness of the study

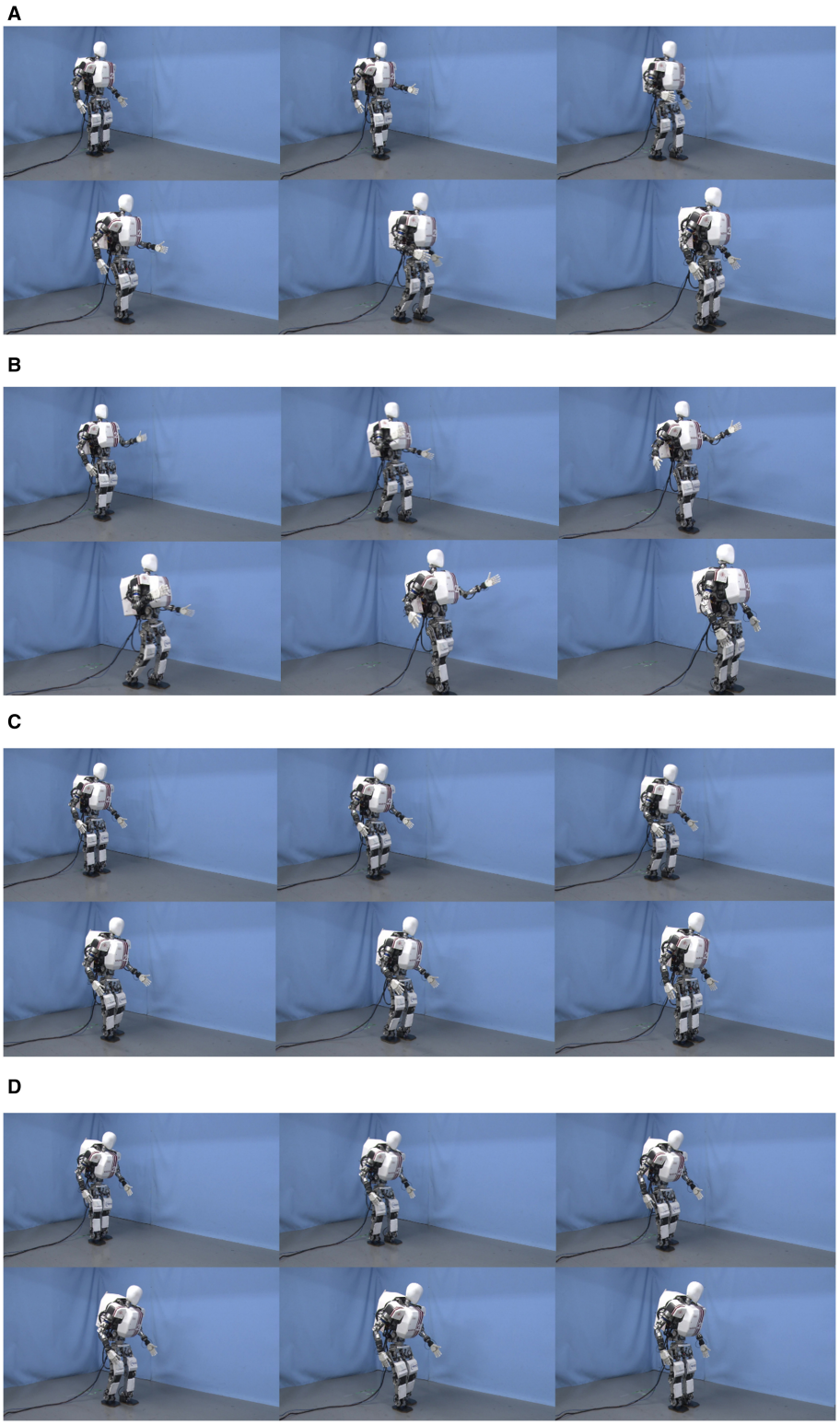


FIGURE 3 | Emotional gait patterns. The following images are captured from the videos shown to the participants. **(A)** represents Happy walk (Natural intensity); **(B)** Happy walk (Exaggerated intensity); **(C)** Sadness (Natural intensity); **(D)** Sad walk (Exaggerated intensity).

subject) instead of encoding for different concepts (Ho and MacDorman, 2010). This questionnaire is made of four different pages. The first page shows a video of the real robot walking without emotion (normal walking) and a text stating "This is the normal emotionless walking robot. Please watch it walking. Normal walking (without emotion)" located at the top of the video. The second, third and fourth page have the same layout. First, a video of the robot walking with an emotion is displayed. Then, the participant is asked the two following questions: (i) "What do you think the robot expressed as emotion?"; (ii)"In what kind of environment and place the movements and the emotions of the robot would be the most relevant?"

This questionnaire was conducted online. The videos used as stimuli lasted between 5 and 10 s and were provided with sound (mainly robot actuators were audible). The participants were given the possibility to replay them at will. The participants were asked to rate the robot and its walking. The questionnaire measures three categories: *Perceived Humanness*, *Eeriness*, and *Attractiveness*. *Perceived Humanness* represents the degree of humanity and human-like characteristics in the robot tested. The *Eeriness* describes the feeling of strangeness, disgust, and familiarity occurring at the same time when something seems natural but some details are not quite conform to the expectation. The *Attractiveness* characterizes the level of comfort and physical attraction we might feel by looking at the robot.

Perceived Humanness: What do you think about the movements of the robot?

Artificial	1	2	3	4	5	Natural
Synthetic	1	2	3	4	5	Real
Inanimate	1	2	3	4	5	Living
Human-made	1	2	3	4	5	Humanlike
Mechanical Movement	1	2	3	4	5	Biological Movement
Without Definite Lifespan	1	2	3	4	5	Mortal

Eeriness: What are your feelings about the robot?

Reassuring	1	2	3	4	5	Eerie
Numbing	1	2	3	4	5	Freaky
Ordinary	1	2	3	4	5	Supernatural
Uninspiring	1	2	3	4	5	Spine-tingling
Boring	1	2	3	4	5	Thrilling
Predictable	1	2	3	4	5	Mortal
Bland	1	2	3	4	5	Uncanny
Unemotional	1	2	3	4	5	Hair-raising

Attractiveness: What do you think of the robot's appearance?

Unattractive	1	2	3	4	5	Attractive
Ugly	1	2	3	4	5	Beautiful
Repulsive	1	2	3	4	5	Agreeable
Crude	1	2	3	4	5	Stylish
Messy	1	2	3	4	5	Sleek

All fields were mandatory. A total of two emotional gaits (Happiness and Sadness) and a neutral gait (for reference and stated as neutral gait) were shown randomly to each participant. For a given participant the intensity of the emotional gaits was fixed: Natural (High) or Exaggerated. For the French people, 23 subjects were randomly exposed to the Natural (High) intensity and 24 to the Exaggerated intensity. For Japanese people, 11 subjects were randomly exposed to the Natural (High) intensity and 11 to the Exaggerated intensity. The results regarding the Autism questionnaire are not discussed in this work.

RESULTS

Attitude toward robots

First we want to explore the pre-conceived ideas about robots for the different factors we are considering for our study. All the participants were not only divided by nationality (French or Japanese) but were also divided according to their *attitude toward robots* (positive or negative), their *age* (young age, middle age or old age), their *familiarity with robots* (not familiar or familiar), and their *interest for robots* (not interested or interested) (**Table 1**). The values are on a scale between −3 and +3 (7-Likert scale), with negative values indicating a disagreement with the item and positive values agreeing with it. *Exposure to robots* indicates how many exposures had the participant with robots through media, events, programming, etc. *Robot preference* shows whether the participant prefer people (reported as negative value) or robot (reported as positive value). *Warmness toward robots* points out whether the participant is cold (reported as negative value) or warm toward robots (reported as positive value). *Warmness toward people* reports whether the participant is cold (reported as negative value) or warm toward people (reported as positive value). *Robots' threat* informs about the participant's feelings on whether robots are more threatening than people (reported as negative value) or the inverse (reported as positive value). *Robots are safe* indicates whether the participant feels that robots are threatening (reported as negative value) or safe (reported as positive value) and *People are safe* whether the participant feels that people are threatening (reported as negative value) or safe (reported as positive value). We performed a multi-factorial ANOVA to examine the effects of the different factors (culture, general attitude toward robots, etc.) on the attitudes (exposure to robots, robot preference, etc.).

Culture. Contrary with the common stereotype depicting Japanese people as people quite fond of robots, French participants feel warmer toward robots than Japanese participants [$F_{(1,\ 50)} = 34.966$, $p < 0.001$] and feel also safer with them [$F_{(1,\ 50)} = 11.428$, $p < 0.01$]. Japanese participants tend to find people not safe [$F_{(1,\ 50)} = 47.594$, $p < 0.001$] while French participants are rather moderate. Both cultures prefer, on average, people to robots but trust more robots than people.

Attitude toward robots. The attitude toward the robots is determined by calculating the mean of the "Prefer robots," "Warm toward robot," "Robot are more threatening," "Robots are safe" items. If the value is less than 0, the attitude is negative; and more than 0, the attitude is positive. While positive-minded participants like equally people and robots, negative-minded

Table 1 | Attitudes toward robot and people per factor.

	Exposure to robots	Robot preference	Warmness toward robots	Warmness toward people	Robots' threat	Robots are safe	People are safe
CULTURE							
French	12.6	−1.63	0.87	1.38	−1.66	1.09	−0.15
Japanese	9.91	−1.41	−1.18***	0.78	−1.41	−0.05**	−1.95***
ATTITUDE							
Positive	13.23	−0.64	1.45	1.32	−1.18	1.86	−0.5
Negative	11.04	−2.0***	−0.36***	1.13	−1.77*	0.19***	−0.89
AGE							
Young	11.4	−1.53	−0.07	1.07	−1.33	0.87	−1.17
Middle	13.1	−1.58	0.58	0.97	−1.74	0.48	−0.48
Old	7.75*	−1.63	−0.12	2.5	−1.88	1.13	0.0
FAMILIAR WITH ROBOTS							
Not familiar	6.14	−1.75	−0.54	1.04	−1.64	0.61	−1.07
Familiar	15.56***	−1.44	0.73**	1.29	−1.54	0.8	−0.49
INTEREST IN ROBOTS							
Not interested	4.92	−2.38	−1.08	1.38	−1.61	−0.38	−0.62
Interested	13.32***	−1.38	0.52**	1.14	−1.57	0.98**	−0.75

*The statistical significance is shown in grayed table cells. (***$p < 0.001$; **$p < 0.01$; *$p < 0.05$).*

Table 2 | Emotion recognition.

Emotions	French	Japanese
Happiness	42.6%	59.1%
Sadness	59.2%	68.2%
Average	51.1%	63.3%

participants clearly prefer people to robots [$F_{(1,\ 50)} = 28.614$, $p < 0.001$]. Similarly, positive-minded participants would feel warmer toward robots [$F_{(1,\ 50)} = 11.892$, $p < 0.01$] and also safer with them [$F_{(1,\ 50)} = 16.854$, $p < 0.001$] while negative-minded participants are more reserved and find that robots are much more a threat than positive-minded participants [$F_{(1,\ 50)} = 4.596$, $p < 0.05$].

Age. The age category is divided in three: young (under 30), middle-aged (between 30 and 50) and old (more than 50). Old participants were less exposed to robots than the younger participants [$F_{(2,\ 50)} = 3.505$, $p < 0.05$].

Familiarity. Participants considering themselves familiar with robots have more exposures to them than non-familiar participants [$F_{(1,\ 50)} = 38.982$, $p < 0.001$] and tend also to be warmer toward robots [$F_{(1,\ 50)} = 5.198$, $p < 0.05$].

Interest. Interested participants have more exposures to robots [$F_{(1,\ 50)} = 42.832$, $p < 0.001$], are warmer to robots [$F_{(1,\ 50)} = 10.629$, $p < 0.01$] and find robots rather safe [$F_{(1,\ 50)} = 7.957$, $p < 0.01$] than the non-interested participants.

We performed a Kolmogorov-Smirnov test and found a statistical difference between the warmth felt toward a robot ($M = 0.19$; $SD = 0.89$) or a human ($M = 1.26$; $SD = 0.47$) [$D_{(11)} = 0.75$, $p < 0.01$]. Regarding safety, participants clearly favor robots ($M = 0.78$; $SD = 0.68$) over humans (M = -0.68; $SD = 0.54$) [$D_{(11)} = 0.8333$, $p < 0.001$].

The Uncanny valley

We chose to use the Ho questionnaire which is a modified version of the Godpseed questionnaire to evaluate reactions regarding the Uncanny valley. That questionnaire was designed to test three different groups of items: *Perceived Humanness*, *Eeriness,* and *Attractiveness*, rated from 1 (low) to 5 (high). We tested the whole questionnaire results for reliability: *Perceived Humanness* (Cronbach's α: 0.77), *Eeriness* (Cronbach's α: 0.85), and *Attractiveness* (Cronbach's α: 0.84). Therefore, the questionnaire has a good reliability.

By using the Ho questionnaire, we measure the possible differences existing in the perception of emotional movements and their link to the Uncanny valley phenomenon. We also investigate if several factors would influence the perception such as the emotional intensity, the type of emotions, the culture of the participants, their attitude toward robots, their age, their familiarity, and their interest.

To begin, we analyze the possible difference in the recognition of the emotions between the cultural groups (**Table 2**). We used Pearson's Chi-squared Test in order to compare the two groups. We found no statistical difference between the two groups (French: 51.1%, Japanese: 63.3%, $\chi^2 = 1.4403$, $p > 0.05$). The recognition rate for each emotion was not significantly different between the groups: Happiness (French: 42.6%; Japanese: 59.1%, $\chi^2 = 1.0466$, $p > 0.05$), Sadness (French: 59.2%; Japanese: 68.2%, $\chi^2 = 1.1773$, $p > 0.05$). We confirm that the participants performed above chance level (20%) ($\chi^2 = 103.9135$, $p < 0.000$).

Is it a valley?

According to Ho et al. (Ho and MacDorman, 2010), the *Eeriness* and *Perceived Humanness* can be plotted together to obtain a graph similar to Mori's Uncanny Valley figure. Nonetheless,

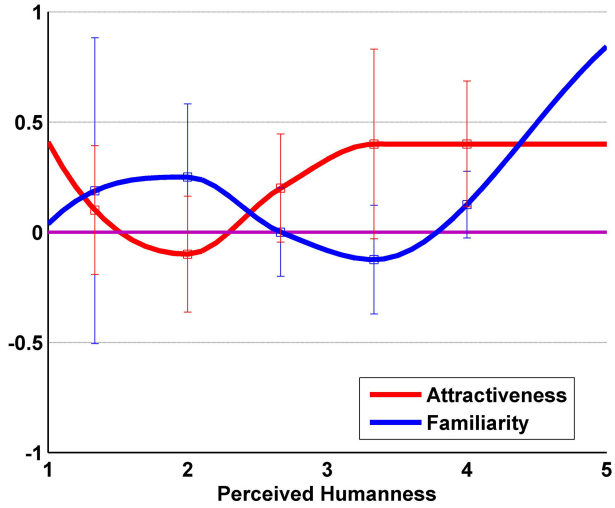


FIGURE 4 | Our Uncanny valley. Perceived Humanness and Familiarity (reverse scaled Eeriness) are plotted in order to describe the Uncanny valley. Attractiveness is also plotted as reference.

the *Eeriness* values have to be transformed into *Familiarity* values by reversing the 5-Likert Type scale (1 becomes 3 and 5 becomes −3) and center the values around 0. We plot *Attractiveness* and *Familiarity* scores against the *Perceived Humanness* score (**Figure 4**). From the plot, we observe two interesting results: an Uncanny valley-like curve for the *Familiarity* score and an another valley-like curve for the *Attractiveness* score.

Cultural difference

As the data we want to analyze are unbalanced and the variables are both categorical and continuous, we use a Generalized Linear Model (GLM) to test each questionnaire item group (*Perceived Humanness, Eeriness,* and *Attractiveness)* with Culture (*French* vs. *Japanese*), Emotion (*Happiness* vs. *Sadness*), and Intensity (*Natural* vs. *Exaggerated*) as independent variables. We found that Intensity is a main effect for *Perceived Humanness* item [$F_{(1,\ 130)} = 11.943$, $p < 0.001$] (Low: 2.42 ± 0.86; High: 2.88 ± 0.69) and the *Nationality* is a main effect for the *Attractiveness* item (*French*: 3.18 ± 0.6; *Japanese*: 2.97 ± 0.5). We tested further the within condition Intensity for French and Japanese participants. The *Intensity* condition only affected the *Attractiveness* felt by Japanese participants [$F_{(1,\ 42)} = 4.172$, $p < 0.05$; *Low*: 3.12 ± 0.5; *High*: 2.81 ± 0.5]. In summary, Japanese people prefer (higher score of *Attractiveness*) Natural Intensity emotions feelings over Exaggerated Intensity emotions and French people prefer neither Natural nor over Exaggerated Intensity emotions.

Attitudes and other factors

We found that the *Attitude toward robots* has a main effect on *Eeriness* and *Attractiveness* questionnaire items. The *Exposure to robots* has also a main effect on the *Attractiveness*. The *Age* × *Attitude* interaction showed to have an significant effect on *Perceived Humanness* and *Attractiveness*. For *Perceived Humanness* we found the following interactions: *Interest* × *Exposures*, *Interest* × *Familiarity*, and *Familiarity* × *Exposures*. For *Attractiveness* we found the following interactions: *Age* × *Familiarity*, *Age* × *Exposures*, *Attitude* × *Exposures*, and *Familiarity* × Attitude. All the statistical results are presented in **Table 3**. To summarize, *Attitude toward Robots* is the main predictor for the *Eeriness* and *Attractiveness* items with the *Exposures to robots* being closely related to it, i.e., if you like robots you will try to be more exposed to them.

Occupation acceptability

We have categorized the occupations in two groups: *Acceptable* and *Non-acceptable* occupation. The *Acceptable* group consists of Police, School, Office, and Hospital related occupations answers and the *Non-acceptable* group of Nowhere answers. We performed an analysis of the correlation using Spearman method to understand how the Occupation acceptability would be influenced by the *Perceived Humanness*, *Eeriness*, and *Attractiveness* felt by the participant. The analysis yielded the following results: *Perceived Humanness* ($r_s = 0.209$, $p < 0.05$) and *Attractiveness* ($r_s = 0.347$, $p < 0.000$). The correlation coefficients suggest that *Attractiveness* is a good predictor (medium effect size) of the Occupation acceptability and *Perceived Humanness* also affects it (small effect size). *Eeriness* ratings did not affect the participants' view on the robot occupation acceptability (not significant) (**Figure 5**). This means that the perception of the Uncanny valley (*Eeriness*) does not affect the acceptability of the robot for a given job and the perceived *Attractiveness* will mostly affect its acceptance, followed by *Perceived Humanness*.

DISCUSSION

First we wanted to investigate factors (cultural background, attitude toward robots, age, interest in robots and familiarity with them), and how those factors might influence our perception of the robot and the Uncanny valley. Then, from those observations, we analyzed their effects on the Uncanny valley phenomenon. Finally, we examined the impact of the participants' perception on the acceptability of the robot's possible occupation.

ATTITUDE TOWARD ROBOTS

Culture

Bartneck and MacDorman studied how people view robots (Bartneck et al., 2007a; MacDorman et al., 2009). Bartneck et al. did a cross-cultural study on people's attitude related to robots. US participants were the most positive toward robots while Mexicans were the most negative toward them (the sample size was small, so the results might biased for Mexicans). He remarked that the Japanese were not as fond as the media seems to portray. MacDorman et al. focused his work on the difference between American and Japanese faculty and also provided proofs against the common stereotype about the Japanese craze for robots. They report no tangible difference observed between American and Japanese faculty. Our results, while focusing on French and Japanese people, support the findings of the two previous studies by Bartneck and MacDorman. Compared to Japanese participants, French participants felt warmer to robots and feel also safer with them. A survey (European Commission, 2012) from the European Union about the attitude toward robots of European

Table 3 | Attitudes and other factors influences on the perception of the robot.

	Perceived humanness		Eeriness		Attractiveness	
	F	*p*	*F*	*p*	*F*	*p*
Age	0.2539	0.8	0.9449	0.4	2.0853	0.1
Attitude toward robots	3.3566	0.06	4.3811	0.04*	4.3079	0.04*
Exposures to robots	0.5283	0.7	0.3424	0.8	6.7485	0.000***
Familiarity	0.024	0.9	2.1774	0.1	2.099	0.2
Interest	0.0911	0.8	2.4179	0.1	0.0536	0.8
Age:Attitude	5.8536	0.004**	2.1674	0.1	5.1924	0.007**
Age:Familiarity	0.7261	0.3	1.8856	0.1	6.9477	0.0096**
Age:Exposures	0.9701	0.4	0.6971	0.6	3.2496	0.006**
Age:Interest	0.7509	0.5	1.5539	0.2	2.2157	0.1
Interest:Exposures	3.5172	0.03*	0.2240	0.7	0.8859	0.4
Interest:Familiarity	4.2636	0.04*	0.7385	0.3	0.6397	0.4
Attitude:Exposures	1.4449	0.2	1.0121	0.4	5.6868	0.0003***
Familiarity:Exposures	9.2221	0.003**	0.5512	0.4	0.0617	0.8
Familiarity:Attitude	0.0040	0.9	0.0461	0.8	4.7741	0.03*
Age:Attitude:Exposures	0.1008	0.96	0.6174	0.6	1.4522	0.2

*The statistical significance is shown in grayed table cells. (***$p < 0.001$; **$p < 0.01$; *$p < 0.05$).*

citizens reports that 70% (67% for French citizens) of them have a positive view on robots, which tend to support our findings. Japanese participants tended to find people not safe while French participants were rather moderate in that aspect. Both cultures prefer, on average, people to humans but trust more robots than people for their safety.

Attitude toward robots

The main effects of the participants' attitude were on their preference between robots and people, the warmness of their feelings for robots and the sense of security they would feel with them. More interestingly, negative-minded participants were not rejecting robots and were rather moderated in their feelings when it comes to robots.

Age

Mitzer et al. found that elderly people have a rather positive attitude regarding assistive technology (Mitzner et al., 2010). Scopelliti et al. on the other hand reported contrary results stating that elderly people express mistrust in technology in general (Scopelliti et al., 2005). When focusing on robots, they noticed that young people rated higher positive feelings and elderly people were the most fearful about robots. While we found that old participants (over 50 years old) feel quite warm toward people and trust people more than the other age categories, the age did not seem to affect the participants' view of the robots. Kuo et al. investigated the influence of age on the attitude toward robots and did not find difference between younger people and elderly people (Kuo et al., 2009) which support our results.

Familiarity and interest

Both participants considering themselves familiar with robots or interested in them have more exposures to them and feel warmer to them too than their non-familiar or non-interested counterparts. Furthermore interested participants prefer robots and find robots safer than the non-interested participants. According to the familiarity principle (the more you are exposed to a thing or a person, the more likeable it will appear to you) (Bornstein, 1989), exposure to robots is most likely the cause explaining the positive view of the robot of the familiar and interested participants.

Finally, we notice that over the five different factors there is no extreme reject of robots. There is a general trend among the participants: they feel closer to other people but they feel safer with robots. As humans, we naturally feel close to beings that look like us and behave like us, especially when we have to choose between organic and inorganic beings. We associate robots with order, logic, and efficiency. Usually represented with a lack of intent, we see robots as predictable beings therefore they might appear safer and less prone to errors than other human fellows. This feeling of safety is important to understand if we want to bring more robotic workers in our society (Takayama et al., 2008). For example, if we consider robots working in fully automated factories the feeling of safety is not mandatory as we will not interact with them. However, for robots in contact with people such as security workers, healthcare helpers, education assistants, and so on. this safety feeling might be influential in the acceptance of a robotic worker.

THE UNCANNY VALLEY

The valley

Some researchers suggested that the Uncanny valley is rather a cliff than a valley (Bartneck et al., 2007b). Their conclusions are drawn from the mapping of the Likeability in the Human-likeness space and fit the data with a quadratic curve. The claim for an Uncanny cliff instead of an Uncanny valley seems to be overstretch in this context. We formulated the hypothesis that our

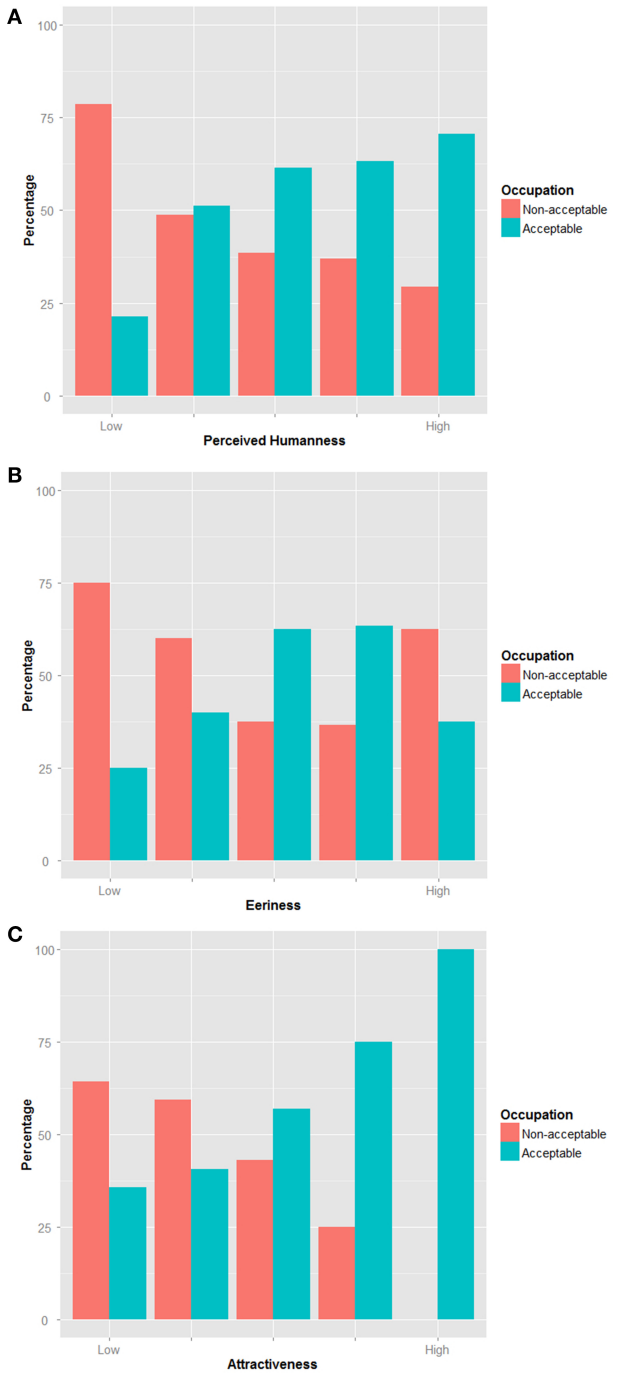


FIGURE 5 | Robot occupation acceptability. The images represent the occupation acceptability of related to the participants' ratings of **(A)** Perceived Humanness; **(B)** Eeriness; **(C)** Attractiveness.

results would describe a valley-like shape similar to the Uncanny valley hypothesis (**Figure 4**). Our results show a similarity with the Uncanny valley hypothesis figure (**Figure 1**) which supports our hypothesis. Tung et al. investigate children's attitude toward robots with two conditions: static and moving (Tung and Chang, 2013). Their results show that the static condition supports well the Uncanny valley static curve. Furthermore, contrary to what is hypothesized in Mori's Uncanny valley hypothesis, the moving condition seems to mitigate the effects of the Uncanny valley instead of amplifying them. In our results, the *Familiarity* values range between a minimum value of −0.12 and a maximum value of 0.84. Those values seem to support the mitigation effect of the motions on the Uncanny valley phenomenon found by Tung et al.

The robot whether perceived totally not human-like (1 on *Perceived Humanness* scale) or fairly human-like to quite human-like (slightly over 3–5 on *Perceived Humanness* scale) is seen similarly attractive to the participants. This would suggest that people would prefer a robot whom they perceive either as quite robot-like or a quite human-like in appearance and behavior. The in-between would be looked down especially if the robot would appear more robot-like than human-like. While we conducted our experiment with only one humanoid robot, we expect this result to be observable in other humanoid robots expressing human-like behavior.

The culture and emotions

We predicted that the culture of the participants will influence in the perception of the *Eeriness* and *Attractiveness*. After analysis, Japanese people were found to prefer (higher score of *Attractiveness*) Natural Intensity emotions feelings over Exaggerated Intensity emotions. Contrary to our expectations, this was the only difference we found regarding the influence of Intensity. This difference may be explained by the Japanese perception of acceptable display of emotions. Several cultures such as Korea and Japan expect neutral display of emotions or low intensity emotions (Trompenaars, 1996). Trompenaars did a cross-cultural study including French and Japanese managers. He found that 42% of the Japanese participants think the emotions should not be displayed overtly and only 14% of French participants think likewise which support our findings. Furthermore, the results from the questionnaire about the participants' attitude toward robots also corroborate that French people have warmer feelings for robots than Japanese people thus explaining the influence of the Culture on the *Attractiveness*.

The attitude and other factors

We predicted that *Attitude toward robots* would influence how people perceive the robot. Positive attitude will rate *Humanness* and *Attractiveness* higher, and *Eeriness* lower and Negative attitude will rate *Humanness* and *Attractiveness* lower, and Eeriness Higher. Our results only confirm that *Attitude toward robots* affects *Eeriness* (Positive: 2.79 ± 0.58; Negative: 2.99 ± 0.54) and *Attractiveness* (Positive: 3.21 ± 0.66; Negative: 3.06 ± 0.54). While the *Age* factor was present in several interactions, it was not by itself a good predictor for any of *Humanness*, *Eeriness*, or *Attractiveness*. In a recent study, MacDorman and Entezari propose to examine nine individual differences (Perfectionism, Neuroticism and Anxiety, Animal Reminder Sensitivity, Personal Distress, Human–Robot Uniqueness, Human–Android Uniqueness, Religious Fundamentalism, and Negative Attitudes Toward Robots) (MacDorman and Entezari, 2015). Alike our findings, they

discovered that Attitudes Toward Robots influenced the sensitivity to the Uncanny valley.

OCCUPATION ACCEPTABILITY

We wondered how our uneasiness might affect our views on a robot having a job and working in contact with us. Our results indicated that only the *Attractiveness* is a good predictor of the Occupation acceptability and *Perceived Humanness* also affects it. The uneasiness felt by the participants did not affect their acceptance of the robot. This result is unexpected as one would think that the uneasiness would lead to rejection regardless of human-likeness and attractiveness of the robot. Here, the "*what is beautiful is good*" stereotype might to overcome the Uncanny effect of the robot and its motions. This stereotype supposes that beauty is strongly related to goodness, therefore good looking persons are better than less attractive persons. This cognitive bias is demonstrated by several studies (Eagly et al., 1991; Agthe et al., 2011). Researchers report that this bias was mostly true when the attractive person to be rated and the evaluator were of opposite sex. When both were of the same sex, the evaluator would feel threatened and then rated lower the attractive person (Agthe et al., 2011). In our study, participants who rated the robot attractive tend to see it working among us, without any effect of its *Eeriness*. Also, since the robot is by design sexless and does not possess any recognizable sexual attribute, positive bias might only apply. This result might be useful for robot designers wanting to overcome Uncanny valley phenomenon. We also found that Perceived Humanness had some influence on the occupation acceptability. Appearance of robots being a predictor on the occupation was studied by Hegel et al. (2009). They found that humanoid robots were thought more fit for occupations similar to humans and animal-like robots were thought adequate as pets or companions. This finding is along the lines of ours: the more the robot would be perceived as human, the more it will be seen fit for work.

CONCLUSION

The Uncanny valley is an intriguing and not well-understood phenomenon. As robotics advances and world population ages, robots will be seen more and more in our daily life. Their behavior, their motions, their emotions might appear alien and thus provoke rejection and uneasiness from users. We propose to study what factors would influence users' impression of the robot. One unique robot, WABIAN-2R, was used for the experiment and only its motions changed, depending on the emotion and the emotional intensity.

One interesting result of this work is that we confirmed that the Uncanny valley to be a highly subjective matter. For the same humanoid robot, some participants perceived it as not quite human and some others found it very human-like. By plotting the participants' reactions to the robot emotional motions, we found the Uncanny valley. Nonetheless, our valley is not much related to the steep depression predicted by Mori when a machine is moving. It was rather a smoother valley similar to his predicted valley describing the still condition. The *Attitude toward robots* was the main influence of the Uncanny valley feeling. Participants who had positive views toward robots rated our robot and its motions less eerie and more attractive than those with negative views.

Another intriguing result is that the perceived *Attractiveness* of the robot had a major effect on its occupation acceptability regardless of how eerie it was rated. Also, the more human like, With a carefully planned external design, it would be possible to minimize any Uncanny valley phenomenon due to strange motions or behavior.

It would be interesting to reproduce the experiment with humanoid robots similar in shape such as ASIMO (Sakagami et al., 2002), HRP-2 (Hirukawa et al., 2004), or even ATLAS from Boston Dynamics. One limitation of the study is the use of videos as stimuli. Videos are useful to understand indirect interaction and impressions from perception. To understand what effect embodiment has, it is necessary to conduct real interaction with users. This will be the next step for our work.

ACKNOWLEDGMENTS

This study was conducted as part of the Research Institute for Science and Engineering, Waseda University, and as part of the humanoid project at the Humanoid Robotics Institute, Waseda University. It was supported in part by JSPS KAKENHI (#26540137 and #26870639) and by Waseda Special Research Funds (#2013A-888). It was also partially supported by SolidWorks Japan K.K and DYDEN Corporation whom we thank for their financial and technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpsyg.2015.00204/abstract

REFERENCES

Agthe, M., Sporrle, M., and Maner, J. (2011). Does being attractive always help? Positive and negative effects of attractiveness on social decision making. *Pers. Soc. Psychol. Bull.* 37, 1042–1054. doi: 10.1177/0146167211410355

Allison, C., Auyeung, B., and Baron-Cohen, S. (2012). Toward brief "Red Flags" for autism screening: the short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls. *J. Am. Acad. Child Adolesc. Psychiatry* 51, 202–212. doi: 10.1016/j.jaac.2011.11.003

Bartneck, C., Croft, E., Kulic, D., and Zoghbi, S. (2009b). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Rob.* 1, 71–81. doi: 10.1007/s12369-008-0001-3

Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. (2007b). "Is the uncanny valley an uncanny cliff?"in *Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2007* (Jeju), 368–373.

Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. (2009a). "My robotic doppelgänger-A critical look at the uncanny valley," in *The 18th IEEE International Symposium on, Robot and Human Interactive Communication, 2009. RO-MAN 2009* (Toyama: IEEE), 269–276.

Bartneck, C., Suzuki, T., Kanda, T., and Nomura, T. (2007a). The influence of people's culture and prior experiences with aibo on their attitude towards robots. *AI Soc.* 21, 217–230. doi: 10.1007/s00146-006-0052-7

Berque, A. (1997). *Japan: Nature, Artifice and Japanese Culture*. Northamptonshire: Pilkington Press.

Bornstein, R. F. (1989). Exposure and affect: overview and meta-analysis of research, 1968–1987. *Psychol. Bull.* 106, 265–289. doi: 10.1037/0033-2909.106.2.265

Dasborough, M. T., and Ashkanasy, N. M. (2002). Emotion and attribution of intentionality in leader–member relationships. *Leadersh. Q.* 13, 615–634. doi: 10.1016/S1048-9843(02)00147-9

Destephe, M., Henning, A., Zecca, M., Hashimoto, K., and Takanishi, A. (2013b). "Perception of emotion and emotional intensity in humanoid robots Gait," in *IEEE Robotics and Biomimetics 2013 (RoBio2013)* (Shenzhen), 1276–1281.

Destephe, M., Maruyama, T., Zecca, M., Hashimoto, K., and Takanishi, A. (2013a). "The influences of emotional intensity for happiness and sadness on walking," in *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2013)* (Osaka), 7452–7455.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., and Longo, L. C. (1991). What is beautiful is good, but…: a meta-analytic review of research on the physical attractiveness stereotype. *Psychol.Bull.* 110, 109. doi: 10.1037/0033-2909.110.1.109

Earhart, H. B. (1982). *Japanese Religion: Unity and Diversity, The Religious Life of Man Series*. Belmont, CA: Wadsworth Publishing Company.

European Commission, (2012). *DG INFSO, Special Eurobarometer 382, Public Attitudes Towards Robots*. Available online at: http://ec.europa.eu/public_opinion/archives/ebs/ebs_382_en.pdf.

Hegel, F., Lohse, M., and Wrede, B. (2009). "Effects of visual appearance on the attribution of applications in social robotics," in *The 18th IEEE International Symposium on Robot and Human Interactive Communication, 2009 RO-MAN 2009* (Toyama: IEEE), 64–71.

Hirukawa, H., Kanehiro, F., Kaneko, K., Kajita, S., Fujiwara, K., Kawai, Y. (2004). Humanoid robotics platforms developed in HRP. *Rob. Auton. Syst.* 48, 165–175. doi: 10.1016/j.robot.2004.07.007

Ho, C. C., and MacDorman, K. F. (2010). Revisiting The Uncanny valley hypothesis: developing and validating an alternative to the Godspeed indices. *Comput. Hum. Behav.* 26, 1508–1518. doi: 10.1016/j.chb.2010.05.015

Ho, C. C., MacDorman, K. F., and Pramono, Z. D. (2008). "Human emotion and the uncanny valley: a GLM, MDS, and isomap analysis of robot video ratings, in *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (Amsterdam: ACM), 169–176.

Hochschild, A. R. (2003). *The Managed Heart: Commercialization of Human Feeling, With a New Afterword*. Berkeley; Los Angeles, CA: University of California Press.

Joosse, M., Lohse, M., Perez, J. G., and Evers, V. (2013). "What you do is who you are: the role of task context in perceived social robot personality," in *IEEE International Conference on Robotics and Automation* (Karlsruhe: ICRA), 2134–2139.

Kaplan, F. (2004). Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots. *Int. J. Humanoid Rob.* 1, 465–480. doi: 10.1142/S0219843604000289

Kuo, I. H., Rabindran, J. M., Broadbent, E., Lee, Y. I., Kerse, N., Stafford, R. M. Q., et al. (2009). "Age and gender factors in user acceptance of healthcare robots," in *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (Toyama), 214–219.

Lubatkin, M. H., Lane, P. J., Collin, S. O., and Very, P. (2005). Origins of corporate governance in the USA, Sweden and France. *Org. Stud.* 26, 867–888. doi: 10.1177/0170840605054602

MacDorman, K. F., and Entezari, S. (2015). *Individual Differences Predict Sensitivity to the Uncanny Valley*. Interaction Studies, IS-D-13-00026R2.

MacDorman, K. F., Minato, T., Shimada, M., Itakura, S., Cowley, S. J., 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).

MacDorman, K. F., Vasudevan, S. K., and Ho, C. C. (2008). Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *AI Soc.* 23, 485–510. doi: 10.1007/s00146-008-0181-2

MacDorman, K. F., Vasudevan, S. K., and Ho, C.-C. (2009). Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *AI Soc.* 23, 485–510. doi: 10.1007/s00146-008-0181-2

MacDorman, K. F. (2006). "Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: an exploration of the uncanny valley," in *ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science* (Vancouver, BC), 26–29.

Mitchell, W. J., Szerszen Sr, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., and MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *iPerception* 2:10. doi: 10.1068/i0415

Mitzner, T. L., Boron, J. B., Fausset, C. B., Adams, A. E., Charness, N., Czaja, S. J., et al. (2010). Older adults talk technology: technology usage and attitudes. *Comput. Hum. Behav.* 26, 1710–1721. doi: 10.1016/j.chb.2010.06.020

Mori, M. (1970). The uncanny valley. *Energy* 7, 33–35.

Ogura, Y., Shimomura, K., Kondo, H., Morishima, A., Okubo, T., Momoki, S., et al. (2006). "Human-like walking with knee stretched, heel-contact and toe-off motion by a humanoid robot," in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (Beijing), 3976–3981.

Pollick, F. E. (2010). In search of the uncanny valley. *User Cent. Med.* 40, 69–78. doi: 10.1007/978-3-642-12630-7_8

Prati, L. M., Douglas, C., Ferris, G. R., Ammeter, A. P., and Buckley, M. R. (2003). Emotional intelligence, leadership effectiveness, and team outcomes. *Int. J. Org. Anal.* 11, 21–40. doi: 10.1108/eb028961

Sakagami, Y., Watanabe, R., Aoyama, C., Matsunaga, S., Higaki, N., and Fujimura, K. (2002). "The intelligent ASIMO: system overview and integration," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2002*, Vol. 3 (Lausanne: IEEE), 2478–2483.

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., and Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* 7, 413–422. doi: 10.1093/scan/nsr025

Scopelliti, M., Giuliani, M. V., and Fornara, F. (2005). Robots in a domestic setting: a psychological approach. *Univ. Access Inf. Soc.* 4, 146–155. doi: 10.1007/s10209-005-0118-1

Takayama, L., Ju, W., and Nass, C. (2008). "Beyond dirty, dangerous and dull: what everyday people think robots should do," in *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (Amsterdam), 25–32.

Thompson, J. C., Trafton, J. G., and McKnight, P. (2011). The perception of humanness from the movements of synthetic agents. *Perception* 40:695. doi: 10.1068/p6900

Trompenaars, F. (1996). Resolving international conflict: culture and business strategy. *Bus. Strategy Rev.* 7, 51–68. doi: 10.1111/j.1467-8616.1996.tb00132.x

Tung, F. W., and Chang, T. Y. (2013). Exploring children's attitudes towards static and moving humanoid robots. *Hum. Comput. Interact.* Part III, 237–245. doi: 10.1007/978-3-642-39265-8_26

Weinshall, T. (1971). Multinational business education-research methodology and attitude study. *Manage. Int. Rev.* 2, 70–87.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 13 June 2014; accepted: 10 February 2015; published online: 25 February 2015.

*Citation: Destephe M, Brandao M, Kishi T, Zecca M, Hashimoto K and Takanishi A (2015) Walking in the uncanny valley: importance of the attractiveness on the acceptance of a robot as a working partner. Front. Psychol. **6**:204. doi: 10.3389/fpsyg.2015.00204*

This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read, for greatest visibility

COLLABORATIVE PEER-REVIEW
Designed to be rigorous – yet also collaborative, fair and constructive

FAST PUBLICATION
Average 85 days from submission to publication (across all journals)

COPYRIGHT TO AUTHORS
No limit to article distribution and re-use

TRANSPARENT
Editors and reviewers acknowledged by name on published articles

SUPPORT
By our Swiss-based editorial team

IMPACT METRICS
Advanced metrics track your article's impact

GLOBAL SPREAD
5'100'000+ monthly article views and downloads

LOOP RESEARCH NETWORK
Our network increases readership for your article

Frontiers
EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on