Categorization difficulty is associated with negative evaluation in the “uncanny valley” phenomenon

Yamada, Yuki
山口大学時間学研究所

Kawabe, Takahiro
NTTコミュニケーション科学基礎研究所

Ihaya, Keiko
九州大学大学院人間環境学研究院

出版情報：Japanese Psychological Research, 2012-05-14. 日本心理学会
バージョン：
権利関係：

Original Article

Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon

Yuki Yamada, Takahiro Kawabe, Keiko Ihaya
(*Three authors contributed equally to this work)

Abstract
Human observers often experience strongly negative impressions of human-like objects falling within a particular range of visual similarity to real humans ('uncanny valley' phenomenon). We hypothesized that negative impressions in the uncanny valley phenomenon are related to a difficulty in object categorization. We produced stimulus images by morphing each two of real, stuffed and cartoon human face images (Experiment 1). Observers were asked to categorize each of these images as either category and evaluated the likability of it. The results revealed that the longest latency, the highest ambiguity in categorization, and the lowest likability score co-occurred at consistent morphing rates. Similar results were obtained even when we employed stimulus images that were created by morphing each two of real, stuffed and cartoon dog images (Experiment 2). However, the effect of categorization difficulty on evaluation was weak when two real human faces were morphed (Experiment 3). These results suggest that the difficulty in categorizing an object as either of dissimilar categories is linked to negative evaluation regardless of whether the object is human-related or not.

This is a preprint version of the article that has been accepted for publication and will be published in Japanese Psychological Research

Correspondence to:
Yuki Yamada, The Research Institute for Time Studies, Yamaguchi University, 1677-1 Yoshida, Yamaguchi, 753-8512, Japan
yamadayuk@gmail.com

Introduction

People often experience an eerie impression or even a sense of revulsion on seeing elaborately designed human-like agents, such as dolls, animated characters, computer-game characters, virtual reality avatars, or robots. Mori (1970)[1] postulated that the likability of a robot gradually increased up to a certain level as its appearance became more human-like, but beyond a certain level of similarity observers suddenly perceived the face as eerie or disgusting. Finally, as the robot's appearance reached the maximum possible degree of similarity, and the perceived strangeness reverted to likability. Mori (1970) referred to this as the "uncanny valley" phenomenon, with reference to this precipitous fall in likability.

Various explanations have been proposed to explain the uncanny valley phenomenon. A previous study suggested that the elicitation of the uncanny valley phenomenon may be related to a fear of mortality, where negative evaluation occurs as a result of an observer doubting their own identity as a living human being on seeing an artificial human-like agent (MacDorman & Ishiguro, 2006). Likewise, an observer may associate disassembled or incomplete human-like agents with scenes of a battlefield containing wounded human bodies after conflict (MacDorman & Ishiguro, 2006). Similarly, MacDorman and Ishiguro attempted to explain the uncanny valley phenomenon in terms of a disgust reaction associated with threat (such as pathogen) avoidance, based on the notion that our cognitive system has evolved a feeling of disgust as a mechanism to avoid infection by harmful bacteria or viruses (Rozin & Fallon, 1987). As illustrated by these hypotheses, researchers have interpreted the uncanny valley phenomenon mainly from an evolutionary point of view. However, it is currently unclear what type of cognitive processing underlies the uncanny valley phenomenon, and how this processing is related to evolutionary factors.

In the present study, we examined a new hypothesis that categorization difficulty is an important factor in the generation of the uncanny valley phenomenon. The uncanny valley occurs when the appearance of a nonhuman agent becomes more human-like (Mori, 1970). As increasing similarity entails the sharing of more visual features (Humphreys, Riddoch, & Quinlan, 1988), it consequently becomes increasingly difficult for observers to categorize agents as human or nonhuman. This difficulty in categorization increases the processing load involved. In addition, previous studies have suggested that relatively high processing fluency is associated with positive impressions (Kuchinke, Trapp, Jacobs, & Leder, 2009; Reber, Schwarz, & Winkielman, 2004; Reber, Winkielman, & Schwarz, 1998). For example, Reber et al. (1998) reported that priming images, which were subliminally presented before target images, increased positive affective judgment on the target images. In turn, it is natural to assume that relatively low processing fluency

[1] Please refer to MacDorman (2005) for English-translated version of Mori (1970).

deteriorates the impression of agents. Therefore, we predict that a decrease in processing fluency due to categorization difficulty may be associated with a negative evaluation of the observed nonhuman agent in the uncanny valley phenomenon. This prediction is consistent with previous theories regarding the relationship between processing fluency and impression formation. However, there have been no studies to directly test the relationship between categorization difficulty and negative impression formation.

A previous study demonstrated that images produced by morphing images of human and computer-generated non-human faces could be used to induce the uncanny valley phenomenon (Seyama & Nagayama, 2007). In the present study, we examined whether the uncanny valley phenomenon was related to the categorization difficulty of morphed images. We used the latency of categorization of morphed images and the point of most ambiguous categorization as indices of categorization difficulty. In addition, we measured the likability ratings of each of the observed images. On the basis of our categorization difficulty hypothesis, we predicted that when the categorization latency and ambiguity were high, likability rating scores would be low. In Experiment 1, we used images produced by morphing images of a real human face, a cartoon human face, and a stuffed human face to elucidate the relationship between categorization difficulty and negative evaluations. In Experiment 2, we used images produced by morphing images of a real dog, a cartoon dog, and a stuffed dog to examine whether the effect of categorization difficulty on object evaluation was specific to stimuli related to the observer's own species. In Experiment 3, we used images produced by morphing images of a male face and a female face and those of two male faces to examine whether categorization difficulty affected the evaluation for average faces.

Experiment 1

This experiment was performed to examine the effect of categorization difficulty on object evaluation. If categorization difficulty affected evaluation, stimuli that were difficult to categorize would induce negative evaluation.

Method

Observers. Twelve observers participated in this experiment. The observers were all naive as to the purpose of this experiment, and all reported that they had normal or corrected-to-normal visual acuity.

Apparatus and Stimuli. Stimuli were presented on a 19-inch CRT monitor (RDF193H; Mitsubishi, Japan) with a resolution of 1024 × 768 pixels, and a refresh rate of 100 Hz. The presentation of stimuli and collection of data were controlled by a computer (Mac Pro; Apple, CA). We used a photometer (3298F; Yokogawa, Japan) to perform gamma correction to linearize the luminance emitted from the monitor.

Stimuli consisted of a fixation point, and morphed images of stuffed human, cartoon human, and real human images (Figure 1). The fixation point was composed of two concentric rings with radii of 0.24° and 0.47° of visual angle, respectively, at a viewing distance of 40 cm. The luminance of each ring was 91.0 cd/m2. We employed grayscale pictures (12.1° × 12.1° of visual angle at a viewing distance of 40 cm) of a stuffed Charlie Brown (*Peanuts*) toy as a stuffed human image and Shinji Ikari (*Neon Genesis Evangelion*) as a cartoon human image. As a real human image, we employed an image produced by morphing two Japanese faces with neutral expressions, selected from a set of face stimuli (JACNeuF: Matsumoto & Ekman, 1988). Each picture was displayed on a gray background (43.5 cd/m2). We generated 11 equally stepped morphed images with morphing proportions ranging from 0% to 100%.

Procedure. The experiment was conducted in a darkened room. The observer's visual field was fixed using a chin-head rest, at a viewing distance of 40 cm. The experiment consisted of two task blocks, one for a categorization task and the other for an evaluation task. In each task, morph pairs were blocked by 11 images for each categorical pair. The order of the blocks was counterbalanced across observers.

The observer initiated each trial by pressing the spacebar on a computer keyboard. The fixation point was presented throughout the experiment whenever the image was not on-screen. In each trial of the categorization task, after a delay of a random duration between 800 and 1200 ms a morphed image was presented and remained on the screen until a response was

This figure is permitted to appear only in the original version of the article

Figure 1. Examples of stimuli used in this study. Note that the examples of the male-female condition in Experiment 3 were different from the stimuli that were actually used in the experiment.

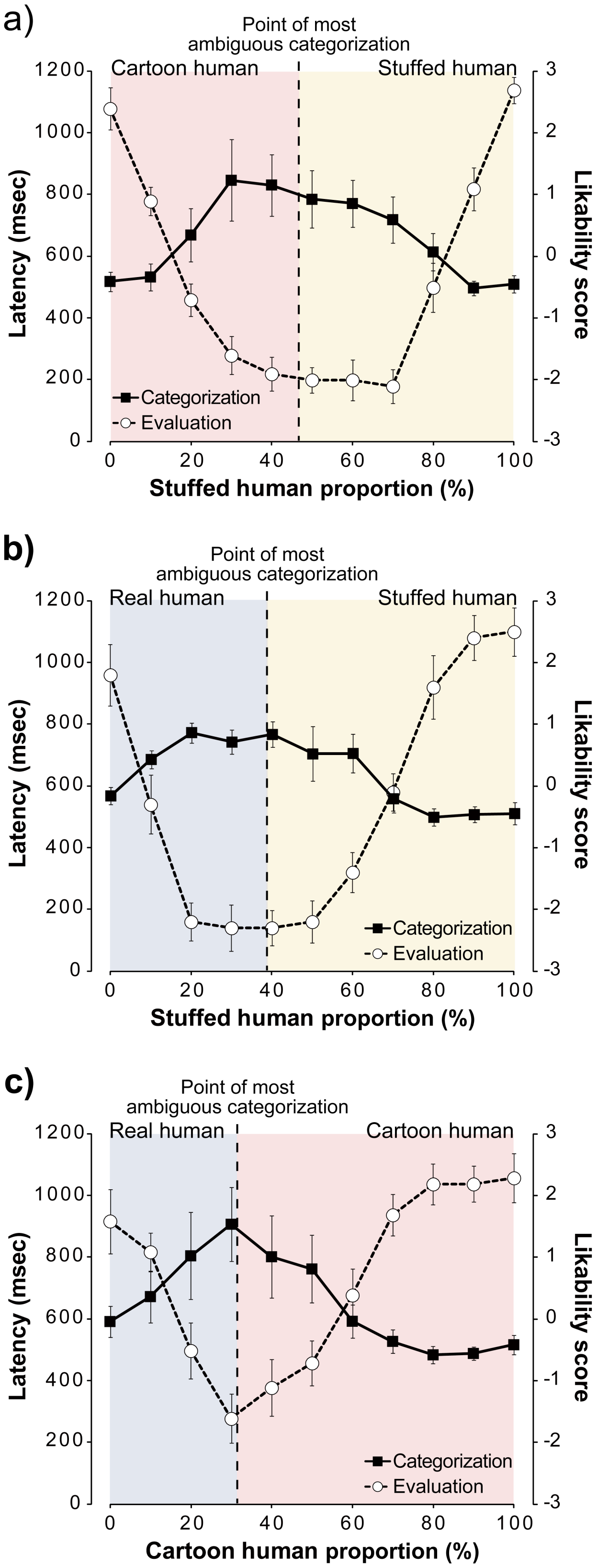

Figure 2. Results in the (A) cartoon-stuffed condition, (B) real-stuffed condition, and (C) real-cartoon condition in Experiment 1. The left and right areas indicated by yellow, red, and blue represent areas in the stuffed, cartoon, and real categories, respectively, centered on the point of most ambiguous categorization. Error bars denote standard errors of the mean.

made. The observer's task was to categorize themorphed image as either category (*e.g.,* real human or cartoon human) by pressing assigned keys as quickly as possible while maintaining accuracy. Each observer performed 330 trials with 3 morphing pairs, 11 images, and 10 repetitions. The trial order was randomized for each observer.

In each trial in the evaluation task, after a delay of 500 ms, a morphed image was presented and remained on the screen until an evaluation was completed. Observers were asked to evaluate the likability of each image using a 7-point scale ranging from –3 (strongly dislike) to 3 (strongly like) using selection keys and a decision key. Rapid responding was not encouraged. Each observer performed 33 trials with 3 morphing pairs and 11 images. The trial order was randomized for each observer.

Results and discussion

The data from two observers were excluded from further analyses because in a post-experiment interview they reported that they evaluated likability of images based on the presence/absence of morphing noise.

The results are shown in Figure 2. A series of one-way analysis of variance (ANOVA) performed on log-transformed response latencies with the morphing proportions as a factor revealed a significant main effect in the cartoon-stuffed [$F(10, 90) = 7.44$, $p < .0001$], real-stuffed [$F(10, 90) = 7.38$, $p < .0001$], and real-cartoon conditions [$F(10, 90) = 6.54$, $p < .0001$]. In the cartoon-stuffed condition, multiple comparisons using Ryan's method (Ryan, 1960) revealed that response latencies for 30% – 60% images were significantly slower than for both 0% and 100% images ($ps < .0006$). In the real-stuffed condition, multiple comparisons indicated that response latencies for 20% and 40% images were significantly slower than for both 0% and 100% images ($ps < .0008$). In the real-cartoon condition, multiple comparisons revealed that response latency for 30% image was significantly slower than for both 0% and 100% images ($ps < .0003$).

A series of one-way ANOVA also revealed a significant main effect of likability scores in the cartoon-stuffed [$F(10, 90) = 39.44$, $p < .0001$], real-stuffed [$F(10, 90) = 41.84$, $p < .0001$], and real-cartoon conditions [$F(10, 90) = 13.22$, $p < .0001$]. In the cartoon-stuffed condition, multiple comparisons revealed that likability scores for 10% – 90% images were significantly smaller than for both 0% and 100% images ($ps < .003$). In the real-stuffed condition, multiple comparisons revealed that likability scores for 10% – 70% images were significantly smaller than for both 0% and 100% images (*ps* < .0001). In the real-cartoon condition, multiple comparisons revealed that likability scores for 20% – 50% images were significantly smaller than for both 0% and 100% images (*ps* < .0004).

Moreover, we calculated the points of maximum latency and minimum likability scores and their 95% confidence intervals by fitting a Gaussian function to log-transformed mean latencies and sign-inverted mean likability scores, respectively, as a function of morphing proportions. For the cartoon-stuffed condition, the points of maximum latency (47.7% ± 4.6%) and minimum likability score (50.0% ± 4.3%) did not differ significantly ($p > .05$). For the real-stuffed condition, the points of maximum latency (35.0% ± 13.6%) and minimum likability score (38.8% ± 3.2%) did not differ significantly ($p > .05$). Furthermore, for the real-cartoon condition, the points of maximum latency (30.7% ± 23.3%) and minimum likability score (36.4% ± 3.2%) did not differ significantly ($p > .05$). Moreover, we used the psignifit program implemented in MATLAB (Wichmann & Hill, 2001a, 2001b) to calculate the point of most ambiguous categorization by fitting a cumulative Gaussian function to the data. For each observer, we calculated the proportion of the trials in which the image was judged as a stuffed human (in the cartoon-stuffed condition) or a real human (in the real-stuffed and real-cartoon conditions) in the categorization task. We assessed the goodness of fit by calculating the deviance and cumulative probability estimate, and confirmed that data were well fitted by the cumulative Gaussian function ($p < .95$). For the cartoon-stuffed condition, we estimated the 95% confidence interval of the mean point of ambiguous categorization (46.7% ± 9.7%). This did not differ significantly from the points of maximum latency and the minimum likability score ($p > .05$). Likewise, the point of most ambiguous categorization did not differ significantly from the points of maximum latency and the minimum likability score in both the real-stuffed condition (38.5% ± 10.6%) and the real-cartoon condition (31.4% ± 7.4%).

In addition, we conducted correlation analysis to examine whether the overall latencies and overall likability scores were related. The results revealed significant negative correlations between these indices in the cartoon-stuffed condition ($r = -.91$, $p < .0002$), in the real-stuffed condition ($r = -.94$, $p < .0001$), and in the real-cartoon condition ($r = -.96$, $p < .0001$).

Taken together, the results of Experiment 1 supported our categorization difficulty hypothesis in that difficulty of categorization indexed by response latency for categorizing an image was related to negative evaluation of the image. Consistent with our predictions, observers showed slow latency for categorization and reported low likability for categorically ambiguous images. Moreover, the points of the maximum latency, the minimum likability score, and the most ambiguous categorization were coincident with each other and the overall latencies and overall likability scores were significantly correlated.

Experiment 2

In contrast to the previous suggestion that the uncanny valley phenomenon occurred between real human and some other categories (*e.g.,* robots, dolls, animated characters, or computer-game characters), our categorization difficulty hypothesis suggested that the uncanny valley phenomenon would occur even when categories were not related to the observer's own species. The uncanny valley phenomenon in nonhuman categories was demonstrated in a previous study (Steckenfinger, & Ghazanfar, 2009) and in the cartoon-stuffed condition in Experiment 1. However, it has not been examined whether the uncanny valley phenomenon occurred between categories that are not related to the observer's own species. Experiment 2 was performed to test this issue by employing images of dogs. If our categorization difficulty hypothesis was valid, the results of this experiment would also show the effect of categorization difficulty on evaluation.

Method

Observers. Ten observers participated in this experiment. The observers were all naive as to the purpose of this experiment, and all reported that they had normal or corrected-to-normal visual acuity.

Apparatus, Stimuli, and Procedure. This experiment was identical to Experiment 1 except we employed three pairs of morphed images of a real dog, a cartoon dog, and a stuffed dog (Figure 1). These images were grayscale pictures of a real Beagle dog, a cartoon Snoopy (*Peanuts*), and stuffed Snoopy toy, respectively. It should be noted that the character Snoopy is based on a Beagle (Schultz, 2004). In the categorization task, observers were asked to categorize the morphed image as either category (*e.g.,* real dog or cartoon dog) by pressing assigned keys as quickly as possible while maintaining accuracy.

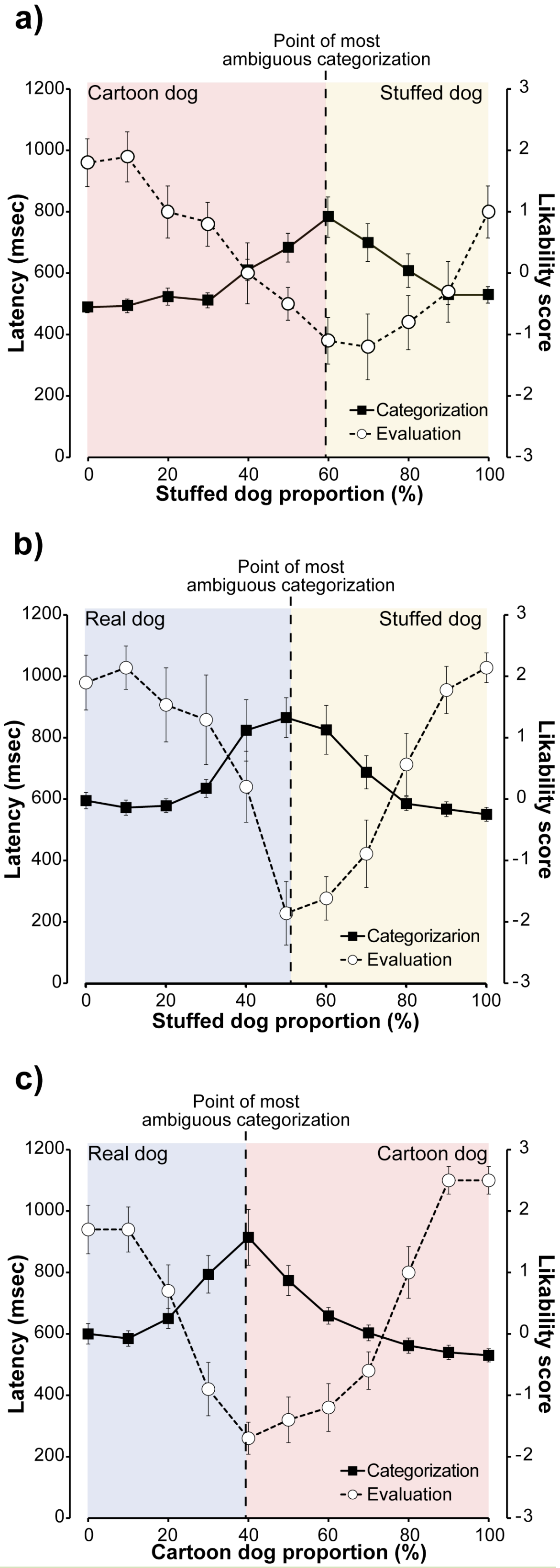

Figure 3. Results in the (A) cartoon-stuffed condition, (B) real-stuffed condition, and (C) real-cartoon condition in Experiment 2. The left and right areas indicated by yellow, red, and blue represent areas in the stuffed, cartoon, and real categories, respectively, centered on the point of most ambiguous categorization. Error bars denote standard errors of the mean.

Results and discussion

The results are shown in Figure 3. A series of a one-way ANOVA revealed a significant main effect of log-transformed latency in the cartoon-stuffed [$F(10, 90) = 18.83$, $p < .0001$], real-stuffed [$F(10, 90) = 17.47$, $p < .0001$], and real-cartoon conditions [$F(10, 90) = 16.85$, $p < .0001$]. In the cartoon-stuffed condition, multiple comparisons revealed that response latencies for 30% – 50% images were significantly slower than for both 0% and 100% images (*ps* < .0001). In the real-stuffed condition, multiple comparisons revealed that response latencies for 40% – 60% images were significantly slower than for both 0% and 100% images (*ps* < .0001). In the real-cartoon condition, multiple comparisons revealed that response latencies for 30% – 50% images were significantly slower than for both 0% and 100% images (*ps* < .0001).

A series of a one-way ANOVA also revealed a significant main effect of likability scores in the cartoon-stuffed [$F(10, 90) = 11.27$, $p < .0001$], real-stuffed [$F(10, 90) = 16.71$, $p < .0001$], and real-cartoon conditions [$F(10, 90) = 24.33$, $p < .0001$]. In the cartoon-stuffed condition, multiple comparisons revealed that likability scores for 20% – 50% images were significantly smaller than for both 0% and 100% images (*ps* < .002). In the real-stuffed condition, multiple comparisons revealed that likability scores for 30% – 60% images were significantly smaller than for both 0% and 100% images (*ps* < .005). In the real-cartoon condition, multiple comparisons revealed that likability scores for 30% – 70% images were significantly smaller than for both 0% and 100% images (*ps* < .0001).

In addition, we calculated the points of maximum latency, minimum likability score, and most ambiguous categorization and their 95% confidence intervals. For the cartoon-stuffed condition, the points of maximum latency (42.5% ± 10.5%), minimum likability score (35.4% ± 4.2%), and most ambiguous categorization (40.6% ± 5.7%) did not differ significantly ($p > .05$). For the real-stuffed condition, the points of maximum latency (49.0% ± 9.2%), minimum likability score (56.4% ± 4.1%), and most ambiguous categorization (51.1% ± 6.9%) did not differ significantly ($p > .05$). Furthermore, for the real-cartoon condition, the points of maximum latency (42.3% ± 10.5%), minimum likability score (47.4% ± 3.0%), and most ambiguous categorization (39.5% ± 5.8%) did not differ significantly ($p > .05$).

Furthermore, significant negative correlations were found between overall latency and overall likability scores in the cartoon-stuffed condition ($r = -.85$, $p < .001$), in the real-stuffed condition ($r = -.87$, $p < .001$), and in the real-cartoon condition ($r = -.83$, $p < .002$).

The results of Experiment 2 supported our prediction based on the categorization difficulty hypothesis. That is, the uncanny valley phenomenon occurred even when the categories that were used in this experiment (*i.e.,* dog) were not related to the observer's own species (*i.e.,* human). Similar to Experiment 1, ANOVA, the analysis of peaks, and the correlation analysis commonly acknowledged the significant effects of categorization difficulty on evaluation.

Experiment 3

Contrary to the categorization difficulty hypothesis, previous studies on facial attractiveness have shown that average (morphed) faces induce positive evaluation (Apicella, Little, & Marlowe, 2007; Langlois & Roggman, 1990; Langlois, Roggman, & Musselman, 1994). The positive evaluation of average faces apparently contradicts the categorization difficulty hypothesis because it is easily expected that the categorization of average faces into either of original (pre-morphing) faces is difficult. In Experiment 3, we addressed this issue using morphing images of a male face and a female face (male-female condition) or those of two male faces (males condition).

Method

Observers. Ten observers participated in this experiment. The observers were all naive as to the purpose of this experiment, and all reported that they had normal or corrected-to-normal visual acuity.

Apparatus, Stimuli, and Procedure. This experiment was identical to Experiment 1 except that we employed two pairs of morphed images of real human faces. These images were grayscale pictures of three male faces and one female face. In the categorization task, observers were asked to categorize the morphed image as either category ("male or female" in the male-female condition and "male A or male B" in the males condition) by pressing assigned keys as quickly as possible while maintaining accuracy.

Results and discussion

The results are shown in Figure 4. A series of a one-way ANOVA revealed a significant main effect of log-transformed latency in the male-female [$F(10, 90) = 13.78$, $p < .0001$] and males conditions [$F(10, 90) = 9.56$, $p < .0001$]. In the male-female condition, multiple comparisons revealed that response latencies for 50% – 60% images were significantly slower than for both 0% and 100% images ($ps < .0001$). In the males condition, multiple comparisons revealed that response latencies for 40% – 60% images were significantly slower than for both 0% and 100% images ($ps < .0004$). A series of a one-way ANOVA also revealed a significant main effect of likability scores in the male-female [$F(10, 90) = 2.19$, $p < .03$] and males conditions [$F(10, 90) = 4.14$, $p < .0002$]. However, multiple comparisons did not indicate that any images were evaluated significantly lower than both 0% and 100% images. Moreover, correlations between overall latency and overall likability scores were not significant in either the male-female condition ($r = -.56$, $p > .07$) or in the males condition ($r = .42$, $p > .19$).

The results demonstrated that evaluation for morphed images did not decrease, while categorization of the images was significantly difficult. That is, categorization difficulty only

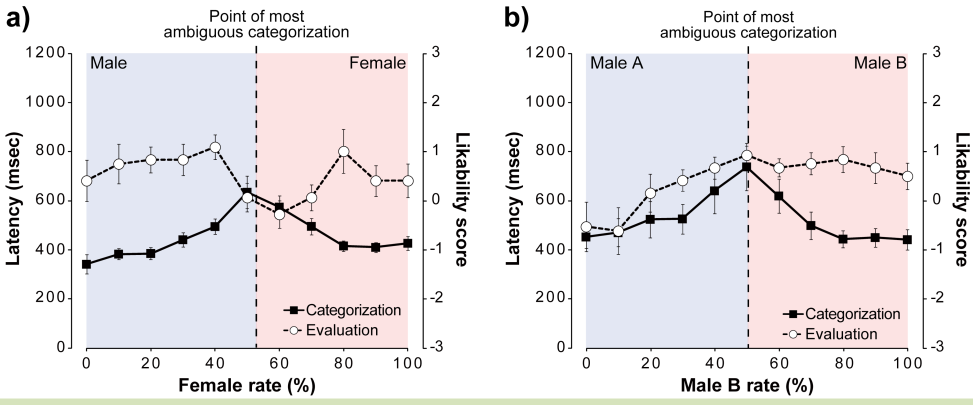

Figure 4. Results of Experiment 3. (A) The left and right areas indicated by blue and red represent areas in the male and female categories in the male-female condition and (B) in the male A and male B categories in the males condition, respectively, centered on the point of most ambiguous categorization. Error bars denote standard errors of the mean.

weakly affected evaluation when two categories involving images to be morphed were real human faces. This issue will be discussed in detail in the General discussion section.

General discussion

The present study was performed to investigate whether the change in visual impression that forms the uncanny valley phenomenon could be explained by the cognitive difficulty involved in categorizing the object. For this purpose, we asked observers to rapidly and accurately categorize morphed images. Moreover, the observers were asked to evaluate the likability of each image. In Experiments 1 and 2, the longest categorization latency, the lowest likability scores, and the highest categorization ambiguity were found to co-occur at consistent morphing proportions, and the latency and likability scores were correlated with each other. However, the effect of categorization difficulty vanished when the categories of stimulus images were both real human faces in Experiment 3.

In accordance with our predictions, the results of Experiments 1 and 2 suggested that the cognitive difficulty involved in the categorization of images is directly related to negative evaluation. Previous studies have shown that higher and lower processing fluencies are related to more positive and more negative evaluation, respectively (Kuchinke et al., 2009; Reber, Schwarz et al., 2004; Reber, Winkielman et al., 2004). In accordance with these previous findings, we propose that processing fluency involving object categorization underlies the modulation of evaluation observed in this study. In Experiments 1 and 2, morphed images with a moderate morphing proportion resulted in ambiguous categorization. Increasing ambiguity would be expected to increase cognitive load, and consequently decrease processing fluency. The reduction of processing fluency may lead to negative evaluations. Thus, the present study provided a new explanation for the uncanny valley phenomenon: *i.e.,* the uncanny valley phenomenon is determined not by visual similarity to a real human (Mori, 1970), but categorization difficulty for an object.

Specifically, the results of Experiment 2 suggested that the uncanny valley phenomenon can occur even when human observers evaluate images of another species. The results of Experiment 2 support the suggestion of Ramey (2005) that the uncanny valley phenomenon is not limited to humanoid robotics. Moreover, Hanson (2006) suggested that the eerie impression for facial images morphed among human-like robots can be decoupled from human realism. On the other hand, the results of Experiment 2 are inconsistent with several previous theories that cannot account for the effects of other-species images, including proposed explanations of the phenomenon as resulting from the detection of deviations from typical human characteristics (MacDorman & Ishiguro, 2006) or the avoidance of inter-human pathological infections (*e.g.,* MacDorman & Ishiguro, 2006). Furthermore, although Ramey (2005) suggested that an uncanny impression is related to the threat of human identity by quantitatively linking qualitatively different categories (*i.e.,* robots vs. human), our findings were inconsistent with this hypothesis in that one of the categories was not necessarily human.

The results of Experiment 3 showed that images categorized with difficulty were not always negatively evaluated, putting limitation on the categorization difficulty hypothesis. The limitation factor is possibly the similarity of categories involving images to be morphed. As shown in Figure 4, we found that changes in likability scores as a function of morphing proportions plotted a concave curve in the male-female condition, as observed in Experiments 1 and 2, but a linear function in the males condition. The former male-female pair was categorically more dissimilar than the latter male-male pair. Both males and females belonged to the human category. On the other hand, cartoon, stuffed, and real human (or dog) categories employed in Experiments 1 and 2 were strongly dissimilar from each other. Considering these categorical similarities and the results of the present experiments, it is possible that the effect of categorization difficulty is prominent when paired categories are dissimilar. To assess the validity of our interpretation, we examined the relationship between categorization difficulty and likability scores for each image used in this study (Figure 5). Supporting our suggestion, a significant correlation was seen in the dissimilar category group (images used in Experiments 1 and 2; $r = -.83$, $p < .0001$), but no such correlation was found in the similar group (images used in Experiment 3; $r = -.08$, $p > .73$).

As an alternative account of the results of Experiment 3, it could be argued that the effect of categorization difficulty occurred regardless of categorical pair, whereas facial attractiveness for average (morphed) faces was enhanced only when real human faces were used (*i.e.,* the male-female and males conditions in Experiment 3), and hence these effects canceled each other out. However, a previous study showed that the effect of averageness on facial attractiveness occurred only when composite faces were made by averaging more than 16 individual faces (Langlois & Roggman, 1990). In the present study, however, we morphed only two individual faces. Therefore, it is unlikely that facial averageness moderated the low likability for images with high categorization difficulty in Experiment 3.

The results reported here were inconsistent with those of a previous study by Seyama and Nagayama (2007); although they conducted rating experiments that were similar to the present study, they did not obtain a clear decrement of evaluation of morphed images. The inconsistency may be explained by three methodological differences. First, the rating items differed between the two studies—Seyama and Nagayama measured pleasantness of images, while we measured likability of images, and categorization difficulty may affect pleasantness and likability scores in different ways. Second, Seyama and Nagayama did not conduct a categorization task. It is possible that prior experience of a categorization task for observers engaging in an evaluation task may have caused negative evaluation. To test this point, a series of a mixed ANOVA was performed on likability scores in all the morphing conditions used in this study, with task order (categorization first vs. evaluation first) as a between-participants factor and morphing proportions as a within-participants factor. The results showed neither significant main effects nor significant interactions in any conditions, suggesting that the prior experience of a categorization task could not explain the inconsistency. Third, Seyama and Nagayama used more realistic artificial faces (*e.g.,* doll and computer graphics) than our stimuli used in Experiment 1 as images that were morphed with real human faces. It is possible that their observers would feel that the categorical similarity between the realistic artificial faces and real human faces was higher than our observers felt about our stimuli, leading to a weak effect of categorization difficulty as in our Experiment 3.

Categorization models are mainly distinguished into three types, *i.e.,* rule-, prototype-, and exemplar-based models. In the rule-based model (Ashby & Gott, 1988; Ashby & Townsend, 1986), categorization difficulty is based on the difference between stimuli and decision rules. In the prototype-based model (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), categorization difficulty is based on the difference between stimuli and an abstract prototype that represents a category. In the exemplar-based model (Medin & Schaffer, 1978), categorization difficulty is based on the average difference between stimuli and all exemplars in a category group stored in memory. According to these models, categorization difficulty is related to the difference between stimuli and memorized representations or decision criteria. If such mechanisms underlie categorization, then categorization difficulty becomes high when the difference becomes large, resulting in the generation of negative evaluation.

It remains unclear what types of cognitive mechanism are related to the effect of categorization difficulty in the negative evaluation of an image. We propose that this phenomenon is related to "stranger avoidance," whereby humans tend to avoid strangers who could potentially harm them physically or impair their genetic fitness. When we are unable to

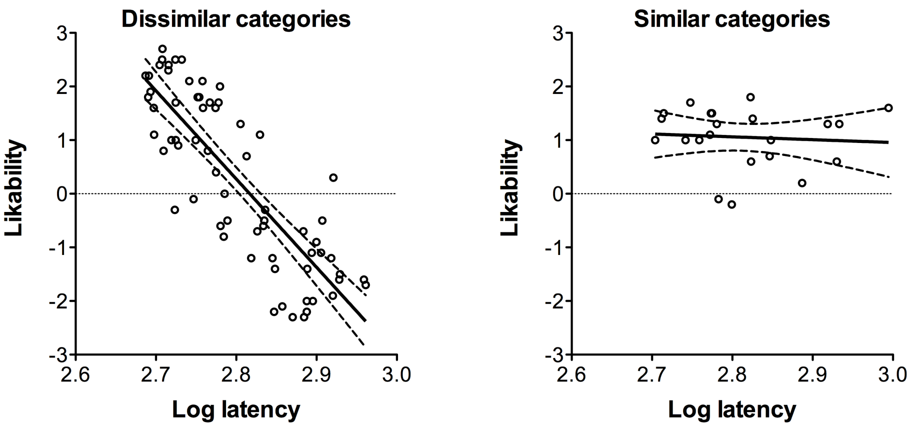

Figure 5. Correlations between categorization difficulty (log-transformed latency) and likability scores for dissimilar category group (left panel) and for similar category group (right panel). The dotted lines denote 95% confidence band.

categorize an object during cognitive processing on categorization discussed above, it is perceived as more strange. In this case, a danger-detection system (Le Doux, 1998; Paradiso, Johnson, Andreasen, O'Leary, Watkins, Ponto, & Hichwa, 1999) may be activated by such uncategorized strangers, leading to avoidance reactions, such as the elicitation of negative impressions. This evolutionarily adaptive function may protect individuals and their genetic information from a large variety of enemies and hazards.

The present results suggest two important issues that should be addressed in future studies. First, it is unclear whether the uncanny valley phenomenon occurs between categories other than animate creatures, such as foods, geometrically differing random figures, and colors. Previous studies on category-specific semantic deficits have suggested that the cognitive processing on categorization for living and non-living objects is dissociated (Forde & Humphreys, 2002). In the present study, we used only stimuli of living objects. Further studies on the effects of categorization difficulty on evaluation of non-living objects will reveal whether the stranger-avoidance system reacts to living strange objects or both living and non-living strange objects. Second, further studies are required to clarify whether the categorization difficulty hypothesis is related to the exaggeration of the uncanny valley in a dynamic display (Ho, MacDorman, & Pramono, 2008; Steckenfinger, & Ghazanfar, 2009; Walters, Syrdal, Dautenhaun, te Boekhorst, Koay, 2008). Mori (1970) proposed that object movement would increase the strength of the uncanny valley phenomenon if the pattern of movement was perceptually strange relative to the movement of real humans. These important issues of the categorization difficulty hypothesis remain unclear. The methods and ideas proposed in the present study will contribute to the resolution of these issues, and the understanding of the cognitive mechanisms underlying the uncanny valley phenomenon.

Acknowledgments

The authors are indebted to Kyoshiro Sasaki for collecting a part of the data.

References

Apicella, C. L., Little, A. C., & Marlowe, F. W. (2007). Facial averageness and attractiveness in an isolated population of hunter-gatherers. *Perception*, **36**, 1813-1820.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **14***,* 33-53.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review,* **93***,* 154-179.

Forde, E. M. E., & Humphreys, G. W. (2002). *Category specificity in brain and mind*. Hove, Sussex: Psychology Press.

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

Ho, C.-C., MacDorman, K. F., & Pramono, Z. A. D. (2008). Human emotion and the uncanny valley: A GLM, MDS, and ISOMAP analysis of robot video ratings. *In ACM/IEEE international conference on human-robot interaction.*

Humphreys, G. W., Riddoch, M. J., Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology*, **5**, 67-103.

Kuchinke, L., Trapp, S., Jacobs, A. M., & Leder, H. (2009). Pupillary responses in art appreciation: effects of aesthetic emotions. *Psychology of Aesthetics, Creativity and the Arts*, **3**, 156-163.

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science*, **1**, 115-121.

Langlois, J. H., Roggman, L. A., & Musselman, L. (1994). What is average and what is not average about attractive faces? *Psychological Science*, **5**, 214-220.

Le Doux, J. (1998). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Touchstone.

MacDorman, K. F. (2005). Androids as experimental apparatus: Why is there an uncanny valley and can we exploit it? *CogSci-2005 Workshop: Toward Social Mechanisms of Android Science* (pp. 108–118). July 25-26, 2005. Stresa, Italy.

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive science research. *Interaction Studies*, **7**, 297-337.

Matsumoto, D., & Ekman, P. (1988). *Japanese and Caucasian facial expressions of emotion and neutral faces (JACFEE and JACNeuF)*. San Francisco: University of California, Human Interaction Laboratory.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207-238.

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

33-35.
Paradiso, S., Johnson, D. L., Andreasen, N. C., O'Leary, D. S., Watkins, G. L., Ponto, L. L., & Hichwa, R. D. (1999). Cerebral blood flow changes associated with attribution of emotional valence to pleasant, unpleasant, and neutral visual stimuli in a PET study of normal subjects. *American Journal of Psychiatry*, **156**, 1618-1629.
Ramey, C.H. (2005). The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. *In Proceedings of the Views of the Uncanny Valley Workshop, IEEE-RAS International Conference on Humanoid Robots*.
Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, **8**, 364-382.
Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science*, **9**, 45-48.
Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, **3**, 382-439.
Rozin, P., & Fallon, A. E. (1987). A perspective on disgust. *Psychological Review*, **94**, 23-41.
Ryan, T. A. (1960). Significance tests for multiple comparison of proportions, variances, and other statistics. *Psychological Bulletin*, **57**, 318-328.
Seyama, J., & Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments*, **16**, 337-351.
Schultz, C. M. (2004). *The Complete Peanuts 1950-1952 (Vol. 1)*. Fantagraphics Books.
Steckenfinger, S. A., & Ghazanfar, A. A. (2009). Monkey visual behavior falls into the uncanny valley. *Proceedings of the National Academy of Sciences of the USA*, **106**, 18362-18366.
Walters, M. L., Syrdal, D. S., Dautenhaun, K., te Boekhorst, R., & 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. *Autonomous Robots*, **24**, 159-178.
Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function. I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, **63**, 1293-1313.
Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function. II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, **63**, 1314-1329.