AMERICAN PSYCHOLOGICAL ASSOCIATION ibnec

Psychology & Neuroscience


ISSN: 1983-3288

2022, Vol. 15, No. 1, 28–42
https://doi.org/10.1037/pne0000281

# Uncanny Valley Hypothesis and Hierarchy of Facial Features in the Human Likeness Continua: An Eye-Tracking Approach

Ivan Bouchardet da Fonseca Grebot[1], Pedro Henrique Pinheiro Cintra[2], Emilly Fátima Ferreira de Lima[1], Michella Vaz de Castro[1], and Rui de Moraes Jr.[1]

[1] Department of Basic Psychological Processes, University of Brasília

[2] Institute of Physics, University of Brasília

***Objective:*** The uncanny valley hypothesis refers to a subjective experience of eeriness to highly human-like objects (e.g., realistic avatars). There is evidence that objects at the human–avatar category boundary along the dimension of human likeness (DHL) are more likely to evoke the uncanny valley effect. Literature has focused on the affective domain of the phenomenon and studies on the cognitive demands are few. Here, we investigate whether perceptual ambiguity could affect the hierarchical processing of facial features. Our study investigated categorical perception of female and male faces along the DHL. ***Method:*** Participants performed a real vs. artificial categorization task and behavioral measures (categorization threshold and response time; RT) were calculated to determine avatar, boundary, and human face conditions. ***Results:*** An analysis on the hierarchy of gaze dwell time in regions of interest (ROI; eyes, nose, and mouth) showed greater dwell time for the nose area of boundary faces compared to the nose area of avatar and human faces. ***Conclusions:*** Results showed that perceptual discrimination difficulty changed the allocation of attentional resources in boundary faces. Such output may contribute on how we process artificial faces and might improve users' experiences from highly realistic characters.

***Public Significance Statement***
The uncanny valley effect is a subjective experience of eeriness to highly human-like avatars. There is evidence that human–avatar ambiguous characters are more likely to evoke the uncanny valley effect. Perceptual discrimination difficulty changes how attention is allocated for human–avatar ambiguous faces. Such result is relevant for researchers interested in how we interact with artificial faces, and for graphics developers concerned on how to improve users' experiences from their highly realistic characters.

*Keywords:* uncanny valley, face perception, gender differences, eye movements, categorization



The uncanny valley hypothesis was originally described by Mori (1970/2012) as the relationship between the resemblance of an object to a human and the emotional response that the object evokes. As an object approaches human resemblance, the emotional response from the observer should be increasingly positive and empathetic. At a certain point, however, the emotional

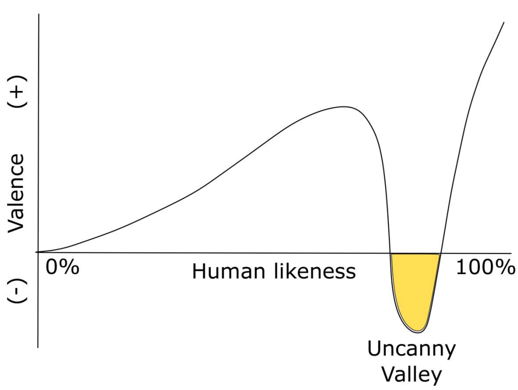
This article was published Online First January 13, 2022.
Ivan Bouchardet da Fonseca Grebot https://orcid.org/0000-0003-4655-2532
Pedro Henrique Pinheiro Cintra https://orcid.org/0000-0001-6604-3597
Emilly Fátima Ferreira de Lima https://orcid.org/0000-0002-4764-1636
Michella Vaz de Castro https://orcid.org/0000-0003-3513-3594
Rui de Moraes Jr. https://orcid.org/0000-0002-9271-9789
We have no known conflict of interest to disclosure.
Correspondence concerning this article should be addressed to Rui de Moraes Jr., Department of Basic Psychological Processes, University of Brasília, Campus Darcy Ribeiro, ICC Sul, sala A1-060, Brasília, DF CEP 70910-900, Brazil. Email: rui.moraes@unb.br

response reverses its valence, and then once again becomes positive (Figure 1). The original hypothesis suggests that a perceptual difficulty in distinguishing between these objects and their correspondent humans evokes negative feelings in the observer, such as eeriness and revulsion (Cheetham et al., 2014; Ho & MacDorman, 2017). The valence of the emotional response is therefore related to the object's degree of realism along a dimension of human likeness (DHL), defined as a linear scale of human likeness (Cheetham et al., 2013).

This phenomenon has been extensively studied in the development of prosthetic limbs, computer animated games and films, and in the development of anthropomorphic robots to prevent characters or objects from falling into the uncanny valley (Fabri et al., 2004; Ho & MacDorman, 2010; MacDorman, Green, et al., 2009; MacDorman, Vasudevan, & Ho, 2009). Research on the uncanny valley is important to understand how human perception reacts to computer graphics and other technological advances, which are increasingly present in our everyday life.

After Mori's 1970 work, different hypotheses on the origin of the uncanny valley effect emerged (Wang et al., 2015). The morbidity hypothesis postulates that highly, but not fully human-like objects (i.e., realistic avatars), may be mistaken for dead bodies, which would evoke an observer's negative response (Mori, 1970/2012). An evolutionary mechanism would play a role in disease avoidance and mate selection (MacDorman, Green, et al., 2009). The predictive coding hypothesis relates to neural computation proprieties of the brain, which is equipped with processes for anticipation and prediction. In case of violation of expectation, that is, when highly human-like agents are not well-predicted by neural model of sensory states, brain activity will be higher, and the perceiver will fell eeriness (Friston, 2010; Moore, 2012). Recent electrophysiological evidence on biological motion supports this hypothesis and states that neural processing in the perception of others is predictive in nature (Urgen et al., 2018). The perceptual mismatch hypothesis assumes that a negative valence peak would be caused by inconsistencies in the realism level of an object's features (e.g., artificial eyes on an almost human-like face) or due to the presence of atypical features in humans (e.g., large eyes; Pollick, 2010). Finally, the categorization ambiguity (or categorical perception) hypothesis proposes that there is ambiguity in categorizing highly realistic artificial objects or characters as human or nonhuman, which would imply a higher perceptual discrimination difficulty for stimuli that fall into the uncanny valley (Cheetham et al., 2014; Kätsyri et al., 2015). Currently, empirical evidence favors the perceptual mismatch and categorization ambiguity hypotheses (Kätsyri et al., 2015; Strait et al., 2017).

**Figure 1**
*Uncanny Valley Model Proposed by Mori (1970/2012)*




Cheetham et al. (2011, 2013) have shown that the difficulty of a categorical perception task varies along the DHL, possibly due to an increase in cognitive costs (Weis & Wiese, 2017; Wiese & Weis, 2020). In both studies, a sharp difficulty increase was found for the "avatar vs. human" categorization task for morphed faces at or surrounding the category boundary (the point of highest categorization ambiguity in the DHL). According to Mori's original hypothesis and the categorization ambiguity hypothesis, faces in the category boundary along the DHL are more likely to evoke the uncanny valley effect. Since the literature has focused on the affective domain as the response variable, these results are relevant as they explore the perceptual and cognitive processing of stimuli that fall in the uncanny valley.

The mind-eye hypothesis states that eye movements reflect what the mind is processing (Beesley et al., 2019; Just & Carpenter, 1980). Therefore, an eye-tracking approach is useful for the study of category processing in the DHL. Particularly, gaze dwell time (i.e., the total time spent gazing at a certain area) indicates perceptual and cognitive loads when discriminating between similar stimuli (Becker, 2011; Shen et al., 2003). Perceptual ambiguity in categorization tasks increases attentional recruitment and processing time, which is reflected in longer dwell times (Barton et al., 2006; Heekeren et al., 2008).

The hierarchy and saliency of facial features, especially the eyes, nose, and mouth in the visual perception of faces (Barton et al., 2006; Chuk et al., 2017; Fraser & Parker, 1986) has either received attention (e.g., MacDorman, Green, et al., 2009) or been directly addressed in research related to the uncanny valley (Cheetham et al., 2013). Cheetham et al. (2013) designed an avatar–human categorization task for faces presented along the DHL. Results showed greater dwell times for eyes and mouth for boundary faces (i.e., most ambiguous faces in the avatar–human morphing continua), reflecting that perceptual discrimination difficulty changed the relative importance of facial features. These results are relevant for experimental psychologists and neuroscientists interested in how we interact with and process artificial faces, as well as for graphics developers and designers concerned on how to improve users' experiences from their highly realistic characters.

The stimuli set used in the experiment conducted by Cheetham et al. (2013) was comprised only of male faces. Besides robust evidence on women's advantage on face processing in different paradigms (e.g., within task "learning + recognition" and simultaneous perceptual matching; Herlitz & Lovén, 2013; Megreya et al., 2011; Pavlova et al., 2016), and under many conditions (e.g., observer's view, gaze direction, race, and time constraints; Godard & Fiori, 2012; Goodman et al., 2012; Lovén et al., 2011; Rehnman & Herlitz, 2006), the literature also reports a female own-gender bias. Accordingly, behavioral (Sommer et al., 2013), eye movement (Coutrot et al., 2016), and event-related potential studies (Wolff et al., 2014) showed that women are better at recognizing female faces. Lewin and Herlitz (2002) hypothesize that women may have some form of specialization in face recognition for female faces, possibly due to prior knowledge and differences in interest. A male own-gender bias is also reported in the literature, although less consistently (Wolff et al., 2014; Wright & Sladden, 2003).

Specifically on eye movements research, previous studies showed different viewing strategies in men and women. For instance, Hall et al. (2010) showed that women had greater dwell time and number of fixation than men to recognize emotional faces; women also looked more at the eyes. Heisz et al. (2013) showed that women made more fixations than men, but there were no gender differences in distribution of fixations in the inner facial features. Coutrot et al. (2016) asked 405 participants to look at 40 videos and showed that men look more at the eyes and women had a more exploratory visual scan (shorter fixation, larger saccades, and scattered eye positions); in addition, women watching actresses gazed more at the left eye. Conversely, Sammaknejad et al. (2017) recently showed no significant gender differences of fixations in the regions of interest (ROI), although women had more transitions from other ROI to the eyes. In general, literature advanced more on the investigation of female and male observers than on the investigation of perception of female and male faces. Studies that investigate gender-related visual scan patterns are relevant insofar as people strategies to gaze at a face have diagnostic cues on the gender of the observer and on the face being observed. These cues are handy to tailor gaze-based models to masculine and feminine populations for many purposes, for example, disorder diagnosis (Coutrot et al., 2016).

The present study was based on Cheetham et al. (2013) and was conceived to tackle potential differences in the perception of male and female faces. Our study aimed to investigate possible differences in the categorical perception of female and male faces along the DHL (human–avatar morphing continua) when viewed by female and male participants. Participants performed a real–artificial categorization task while behavioral measures were taken (categorization response, categorization response time [RT], and gaze location). The averages of the categorization thresholds and RTs were used to set the most uncertain morph level to determine boundary human faces. The main analysis verified the hierarchy of gaze dwell times in predetermined ROI (i.e., areas encompassing a face feature: eyes, nose, and mouth) for avatar, boundary and human faces of male and female models. Based on previous experimental findings, we hypothesized greater dwell times for ROI of boundary faces, particularly the eyes and the mouth. In addition, a general advantage—specified here as faster RTs and lower dwell times—in face processing was expected for women, in particular when viewing female faces. In addition, since there is evidence for a more exploratory visual scan in women, it was expected lower differences among ROI.

## Method

### Participants

Twenty-nine university students (15 women; $M_{age}$ = 22.2 years, *SD* = 3.8, range = 18–34 years) took part in the experiment with no previous knowledge of the investigation topic and the study's goals. The participants had normal or corrected-to-normal vision (as assessed by a Snellen chart) and reported no history of ocular, psychiatric, or neurological disorders. Participants were allowed to wear prescription glasses or contact lenses during the experiment. All participants read and signed a statement of consent approved by the Human and Social Sciences Research Ethics Committee of the University of Brasília (CAAE 11946519.8.0000.5540). Participation was voluntary and no compensation was offered.

### Stimuli

Face stimuli were comprised of 12 white models (7 male and 5 female) taken from the Karolinska Directed Emotional Faces (KDEF) database (Lundqvist et al., 1998). For each model, frontal and side photographs were used to create a detailed 3D head using FaceGen Modeller 3.1 software (Singular Inversions Inc., Toronto, Canada). All texture information was then removed from the face to create an artificial avatar face that was devoid of facial details, while keeping geometric consistency with its reference model. Both the original KDEF photographs and the avatar face images were then used to create 10 morph continua images for each of the 12 models. All resulting images were converted to grayscale and cropped to an oval shape using Adobe Photoshop CS6 13.0 (Adobe Inc., San José, USA) to preserve only internal facial features. In total, 120 stimuli were created (10 for each model). Five male and five female models were used as stimuli for the experimental task and the remaining two male models were used for training. Figure 2 (Panel A) illustrates the procedure for stimuli creation. At a viewing distance of 60 cm, the facial stimuli (313 × 407 pixels) subtended a visual angle of 10° × 13°. The stimuli set is available at https://osf.io/eqx7w/.

### Apparatus

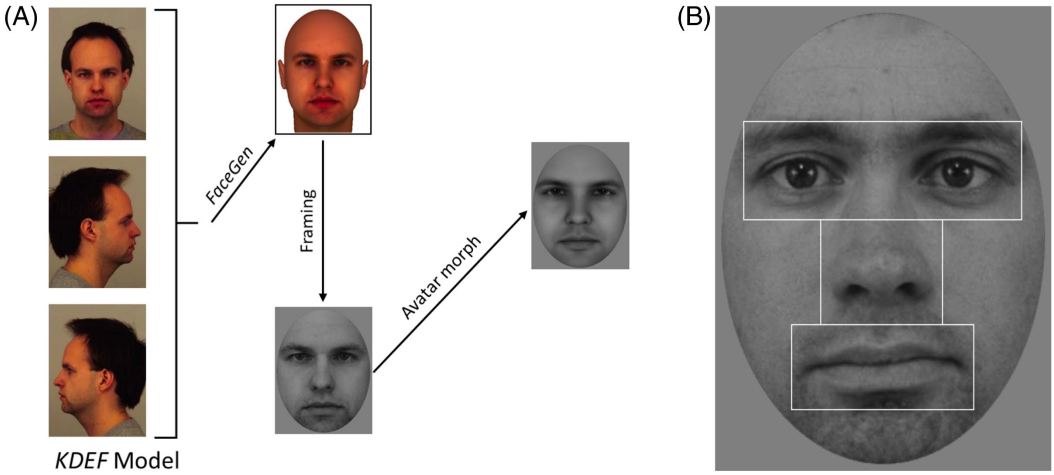
The experiment took place in a sound attenuated room, and stimuli were presented on a 21-in. LCD monitor (60 Hz refresh rate). Eye movement data were recorded using a 200 Hz binocular eye-tracker (Arrington Research Inc., Scottsdale, USA) with a spatial resolution of .15° visual arc and .25° visual arc accuracy. Arrington's proprietary software Viewpoint Eye-Tracker pre-processed and relayed the data to MATLAB version 9.7 (Mathworks Inc., Sherborn, USA). A chin and forehead rest was used to stabilize participants' heads. Psychtoolbox 3 (Brainard, 1997; Kleiner et al., 2007) in MATLAB was used for running the experimental task, displaying the stimulus, and collecting responses.

### Procedure

Participants were tested individually in a single session for approximately 45 min. Both the training and the experimental task followed the same procedure. Participants were given written and illustrated instructions on the computer monitor. After instructions were given, a 16-point random-order grid-pattern calibration procedure took place, followed by 20 randomly ordered training trials. Once training had been completed, the experimental task began, which was comprised of 100 randomly ordered trials (10 trials per morphing level). A 1-min pause was given when the first half of the experiment was completed.

Each trial began with a horizontally centered black fixation point being presented for 1,500 ms either on the left or right side on a screen filled with uniform medium gray. Participants were instructed to gaze at the lateralized fixation point and were video monitored by the experimenter. A centralized oval-cropped grayscale face was then presented, and the participants were instructed to identify each stimulus quickly and accurately as either "real" (human) or "artificial" (avatar) by pressing the A or L keys, respectively, on a standard computer keyboard. When the response

**Figure 2**
*Stimuli Used in the Experiment*




*Note*. Stimuli creation (Panel A) and regions of interest (ROI) sizes: eyes 266 × 88 pixels, nose 116 × 94 pixels, and mouth 174 × 76 pixels (Panel B). KDEF stimuli IDs: AM02NES (Panel A) and AM04NES (Panel B), with permission from Karolinska Institute.

was given, or after the maximum RT of 3,000 ms had been reached, the initial fixation-point screen was presented again starting the subsequent trial.

## Results

### Forced-Choice Categorization

In order to define the most uncertain morph level associated with the uncanny valley effect, we calculated the percentage of faces categorized as real (i.e., human) along the DHL for each participant. We then fitted logistic function models at the individual level. The models were not conditioned to a fixed maximum value for the curve (i.e., height). A sigmoid-shape function was observed for all participants and the data showed high adjustment, as determined by the value of $R^2$ (range = .69–1.00, $M$ = .95, $SE$ = .18). Raw and processed data are available at https://osf.io/eqx7w/. The following analysis on the forced-choice categorization considered two parameters derived from the individual fitted curves: the categorization threshold (the curve midpoint, where the probability to categorize a face as real is equal to 50%) and the curve slope. An in-house code written in Python fitted the curves and calculated the psychometric function parameters (Cintra, 2020). All the statistical analyses were conducted using jamovi 1.6 (The jamovi Project, 2020). We used Greenhouse-Geisser correction when sphericity was violated and Bonferroni adjustment for multiple comparisons. The significance level was set at 5%.

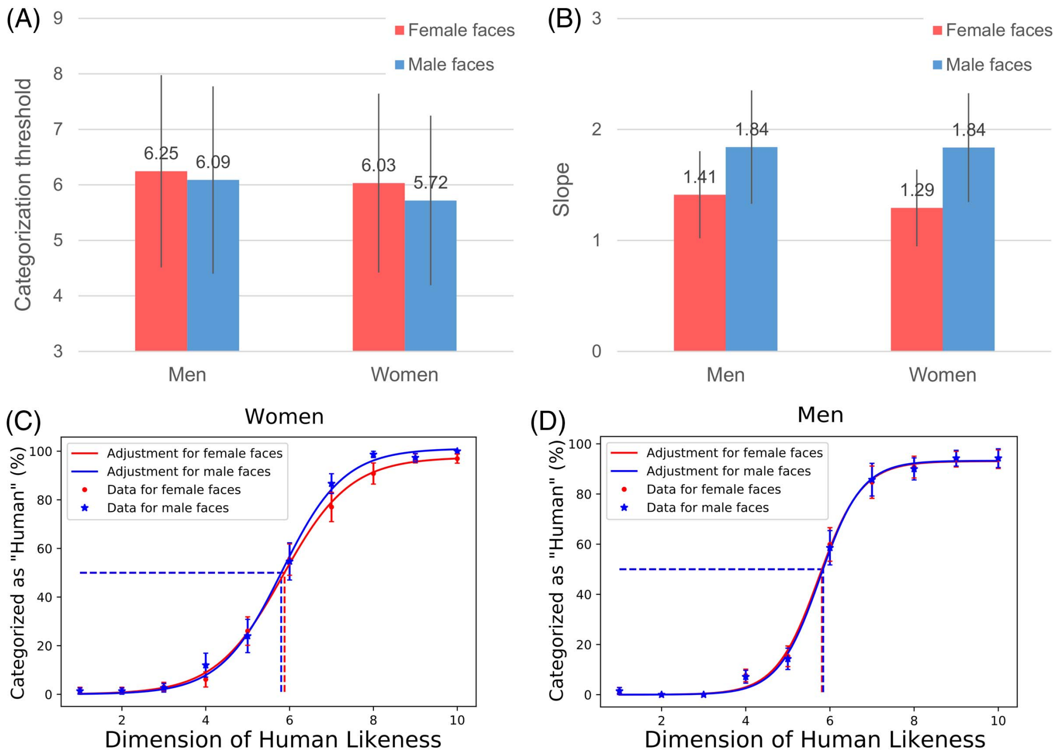
#### *Categorization Threshold*

The categorization threshold indicates the DHL level that corresponds to the ordinate point associated with the observer's maximum uncertainty (50%) in the avatar–human categorization task. A two-way repeated-measures ANOVA was performed with face gender (female and male facial stimuli) as the within-participant factor and participant gender (men and women) as the between-participant factor for the categorization threshold. Figure 3, Panel A, shows the mean threshold along standard errors of the mean for each experimental condition and group; these means range from 5.72 to 6.25 in the DHL. Analysis of variance (ANOVA) did not reveal a significant main effect of face gender, $F(1, 27) = 3.20$, $p = .085$, $\eta_p^2 = .11$, and participant gender, $F(1, 27) = 1.22$, $p = .279$, $\eta_p^2 = .04$. The interaction was also nonsignificant, $F(2, 56) = .35$, $p = .560$, $\eta_p^2 = .01$.

Results of the categorization threshold showed that morph level 6 along the DHL is the most ambiguous stimulus for the avatar–human categorization task for both men and women observers when perceiving both male and female faces.

**Figure 3**

*Properties of the Psychometric Function for the Perceptual Categorization Task*




*Note.* Average of categorization threshold (Panel A) and slope (Panel B) of the individual curves, and logistic regression curves from pooled data for women (Panel C) and men (Panel B). Bars indicate the standard error (Panels A and B) and adjustment error (Panels C and D).

Thus, the faces in the 6th level of our morphing continua would best represent those which fall into the Uncanny Valley (according to the categorization ambiguity framework).

### *Slope*

The slope derived from the individual fitted logistic curves indicates categorical perception for the faces presented in the task. That is, despite a linear continuum in the DHL, the psychometric functions obtained in our task indicate a cognitive process of categorization as observed by the S shaped curves obtained. In order to control extreme values, slopes > 4 were replaced by the mean of the condition[1] (6 values were replaced for both female and male face conditions). A two-way repeated-measures ANOVA was performed with face gender (female and male facial stimuli) as the within-participant factor, and participant gender (men and women) as the between-participant factor for the categorization threshold.

Figure 3, Panel B, shows the mean slope values along standard errors of the mean for each experimental condition and group; these means range from 1.29 to 1.84. ANOVA revealed a significant main effect for face gender, $F(1, 27) = 15.02, p < .001, \eta_p^2 = .36$. The slope was sharper for categorizing male faces (1.84) when compared to female faces (1.35). No main effect for participant gender was found, $F(1, 27) = .21, p = .653, \eta_p^2 = .01$. The two-factor interaction was also nonsignificant, $F(1, 27) = .19, p = .666, \eta_p^2 = .01$.

[1] The cutoff value (4) was specified after observation of the frequency distribution for slope values in our participants sample. The 95% interval of our frequency distribution had a maximum value of 4.1 (considering the range from 0 to α, where α is the threshold representing 95% of area under the curve).

Results on the slope showed that participants presented categorical perception in the experimental task. In addition, a sharper avatar–human categorization seems to occur for male faces regardless of participant gender. However, such output must be interpreted with care given the substantial number of values replaced.

### *Psychometric Function Proprieties of the Pooled Data*

In addition to data fitting at the individual level, we built curves from the pooled data. Figure 3 shows the adjustment for female and male faces perceived by women (Panel C) and men (Panel D). Regarding the categorization threshold (and its adjustment error), men had similar values for male (5.84 ± .94) and female faces (5.81 ± .91). Women also had similar categorization threshold values for male (5.87 ± .56) and female face (5.80 ± .35) conditions. However, men had a higher slope value (female face = 1.96 ± .16; male face = 2.02 ± .16) than women (female face = 1.24 ± .06; male face = 1.36 ± .10) regardless of the gender of the facial stimulus.

Results on the pooled categorization thresholds show a slight difference from the threshold averaged from the individual curves. Such differences are within the boundaries of the threshold adjustment error. Both the averaged and the pooled-data threshold show that faces presented in the 6th level of our morphing continua would best represent the uncanny valley faces. Therefore, the DHL will be represented by three discrete categories in the subsequent analyses: avatar faces (morphing levels: 1–2), boundary faces (morphing level: 6), and human faces (morphing levels: 9–10).

The agreement between the averaged and the pooled data for the categorization threshold was not observed for the slope. The averaged data showed greater slope values for male faces compared to female faces regardless of participant gender. On the other hand, the pooled data showed greater slope values for men regardless of facial stimuli gender. Therefore, an additional Bayesian analysis was conducted (Turner & Van Zandt, 2012) to calculate the slope that best fit to describe such parameter. Since the slope value is a subsidiary measure, this analysis was presented as Supplemental Material found at https://osf.io/eqx7w/. It revealed that the slope is higher for men and that male faces showed larger deviations from the "common" value. The Bayesian method also supports that the slope value calculated from the pool of data better describes the "real slope" value of our data. Despite the inconsistency, both averaged and pooled slope values endorsed a categorical perception.

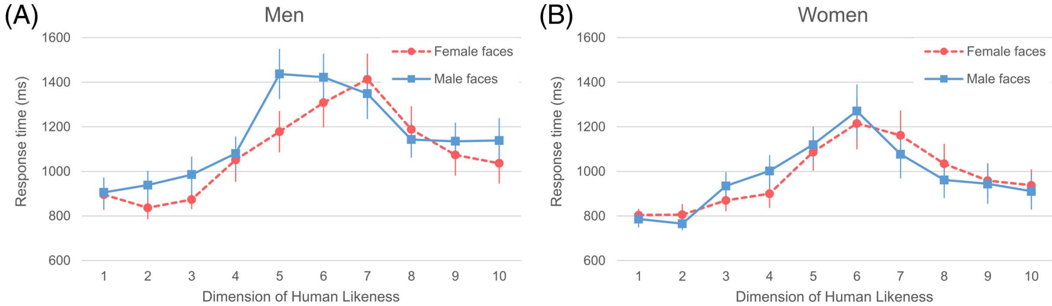
## Response Time

Previous studies showed that a bell-shaped curve is observed for RTs along the DHL, peaking at or near the categorization-threshold morphing level (Cheetham et al., 2011, 2013). Higher RTs are associated with difficulty and complexity in cognitive processing. Thus, long latencies surrounding the categorization threshold support this area of the DHL as the area of the greatest categorization ambiguity. Figure 4 shows the mean RT along standard errors of the mean for each DHL level for men (Panel A) and women (Panel B) when categorizing female and male faces.

A three-way repeated-measures ANOVA was performed with DHL (avatar, boundary, and human faces) and face gender (female and male facial stimuli) as within-participant factors and participant gender (men and women) as the between-participant factor for the RT. As predicted, ANOVA revealed a significant main effect of DHL, $F(2, 54) = 37.59$, $p < .001$, $\eta_p^2 = .58$. Participants were slower to categorize boundary faces (1,322 ms) than human (1,029 ms; $p = .006$) and avatar faces (850 ms; $p < .001$), which in turn had a significant mean difference between them ($p < .001$). No main effect for face gender, $F(1, 27) = 2.14$, $p = .155$, $\eta_p^2 = .07$, and participant gender was found, $F(1, 27) = 1.95$, $p = .174$, $\eta_p^2 = .07$. None of the two-factor interactions were significant: DHL × Participant gender, $F(2, 54) = .134$, $p = .875$, $\eta_p^2 < .01$; Face gender × Participant gender, $F(1, 27) = 1.92$, $p = .177$, $\eta_p^2 = .07$; and DHL × Face gender, $F(2, 54) = 1.495$, $p = .233$, $\eta_p^2 = .05$. ANOVA also revealed a nonsignificant three-factor interaction among DHL × Face gender × Participant gender, $F(2, 54) = 1.495$, $p = .233$, $\eta_p^2 = .05$.

Graphical and statistical analyses on the RT showed a high processing time cost for boundary faces. The RT data support results on the categorization threshold that faces presented in the 6th level of our morphing continua are the most ambiguous stimuli, and therefore would best

**Figure 4**
*Response Time in the Dimension of Human Likeness*




*Note*. Means of the response time in the dimension of human likeness for the "real vs. artificial" categorization task for female and male faces performed by men (Panel A) and women (Panel B). Bars indicate the standard error.

represent the uncanny valley faces. Results also showed that avatar faces are categorized faster than human faces.

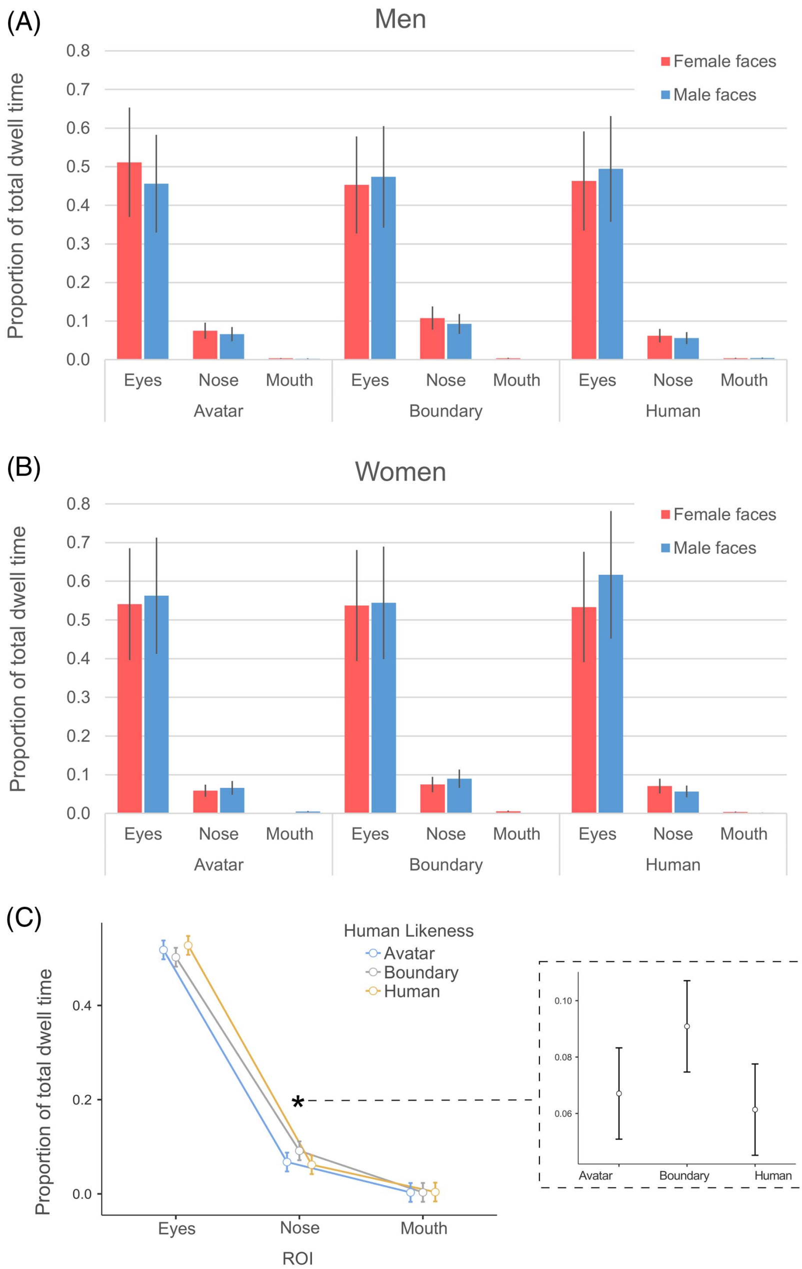
## Eye Movements

Once the most uncertain morph level was established by categorization threshold and RT response variables, subsequent analyses focused on the main goal of the study: to verify the hierarchy of facial features (eyes, nose, and mouth) for avatar, boundary, and human faces. These features were delimited by predetermined ROI (see Figure 2, Panel B). Areas of the ROI remained constant for all stimuli. To check for hierarchical processing of ROI, we computed dwell times based on the results obtained by Cheetham et al. (2013). We only settle the total dwell time to test our hypotheses to avoid analyzing multiple metrics (i.e., data fishing; see Orquin & Holmqvist, 2018). In their study, the dwell time was defined as the "proportion of total fixation duration" within each ROI. However, the definition of fixation is parameter-dependent (gaze position for a minimum of *X* ms inside an area of *Y* degrees of visual angle) and the definition itself raises conceptual confusion in the area (see Hessels et al., 2018, for a discussion). Thus, depending on definition and parameters adopted, dwell time results may vary. In an attempt to circumvent this limitation, we opted for a simpler measure of the total "dwell time": the proportion of the *XY* coordinates of gaze data recorded at the eye tracker sample rate (i.e., capture rate) during the stimuli presentation. The dwell time in each ROI was averaged for each participant, for male and female faces in the DHL conditions for further statistical analysis. Gaze data recorded on the entire screen was used to calculate the dwell time, which is a different approach from Cheetham et al. (2013) that only considered coordinates inside the face area.

### *General ROI Dwell Time*

Figure 5 shows the mean dwell times along standard errors of the mean for each ROI for female and male faces in the DHL conditions for men (Panel A) and women (Panel B) in the real–artificial categorization task. An exploratory omnibus four-way repeated-measures ANOVA was performed for an initial general analysis with ROI (eyes, nose, and mouth), DHL (avatar, boundary, and human faces), and face gender (female and male facial stimuli) as within-participant factors and participant gender (men and women) as a between-participant factor for dwell time. Results on this general analysis of ROI showed a hierarchical processing, that is, participants gazed more at the eyes, nose, and mouth, respectively. The complete results are available at https://osf.io/eqx7w/ as Supplemental Material 1. Since: (1) this hierarchical pattern in the proportion of dwell time was expected (Barton et al., 2006) and (2) total dwell time is not appropriate to infer conclusions about one ROI receiving more attention than another ROI (Orquin & Holmqvist, 2018), independent ANOVAs for each ROI were previously planned.

**Figure 5**
*Dwell Time in Facial Regions of Interest*




*Note*. Dwell time means (i.e., averages of the proportion of total gaze duration) in facial regions of interest (ROI; eyes, nose, and mouth) for the "real vs. artificial" categorization task for female and male faces performed by men (Panel A), women (Panel B), and the total sample (Panel C). Bars indicate the standard error.

Considering the low proportion of dwell time on the mouth[2], we only ran ANOVAs for the eyes and the nose ROI.

### *Eye Dwell Time*

A three-way repeated-measures ANOVA was performed with DHL and face gender as within-participant factors, and participant gender as a between-participant factor for the dwell time in the eyes area. ANOVA did not reveal a main effect for DHL, $F(2, 45) = .87, p = .406, \eta_p^2 = .03$, face gender, $F(1, 27) = 1.90, p = .179, \eta_p^2 = .07$, and participant gender, $F(1, 27) = 1.63, p = .213, \eta_p^2 = .06$. All interactions were nonsignificant: DHL × Participant gender, $F(2, 45) = .31, p = .692, \eta_p^2 = .01$; Face gender × Participant gender, $F(1, 27) = 2.16, p = .154, \eta_p^2 = .07$; DHL × Face gender, $F(2, 52) = 2.35, p = .107, \eta_p^2 = .08$; and DHL × Face gender × Participant gender, $F(2, 52) = .97, p = .384, \eta_p^2 = .03$.

Results showed that participants had no differences on dwell time of gaze on the eyes area when categorizing avatar, boundary, and human faces as "artificial" or "real." Neither the face gender nor the participant gender influenced the results.

### *Nose Dwell Time*

The same three-way repeated-measures ANOVA [participant gender (DHL × Face gender)] was then employed for the dwell time in the nose area. ANOVA revealed a significant main effect of DHL, $F(2, 54) = 7.53, p < .001, \eta_p^2 = .22$, indicating that boundary faces had a greater proportion of gaze dwell time on the nose area (.10) than avatar (.07; $p = .014$) and human faces (.06; $p = .002$), which did not differ between them ($p > .999$). There was no significant main effect of face gender, $F(1, 27) = .271, p = .607, \eta_p^2 = .01$, and participant gender, $F(1, 27) = .291, p = .594, \eta_p^2 = .01$. None of the two-factor interactions were significant: DHL × Participant gender, $F(2, 54) = 1.02, p = .367, \eta_p^2 = .04$; Face gender × Participant gender, $F(1, 27) = .79, p = .380, \eta_p^2 = .03$; and DHL × Face gender, $F(2, 54) = .17, p = .843, \eta_p^2 < .01$. The triple interaction was also nonsignificant, $F(2, 54) = .676, p = .513, \eta_p^2 = .01$.

Results showed that participants spent more time gazing at the nose area of ambiguous faces compared to the nose area of avatar and human faces during a real vs. artificial categorization task. As with the eye area, neither the face gender nor the participant gender influenced the results. All statistical analyzes presented in the Results section are available at https://osf.io/eqx7w/.

## Discussion

We aimed to investigate potential differences on categorical perception of female and male faces presented along the DHL. Specifically, we investigated the relative importance of eyes, nose, and mouth ROI inferred by the total dwell time in a real–artificial categorization task. Results showed a hierarchical gaze allocation with a greater dwell time for the eyes, nose, and mouth, respectively. We found no effect of the face gender, participant gender, nor its interaction. However, participants spent more time gazing at the nose of boundary faces (i.e., most ambiguous faces in the avatar–human morphing continua) compared to nose of avatar and human faces. Boundary faces are more likely to fall in the uncanny valley.

Therefore, our study partially replicates the results of Cheetham et al. (2013). Both studies found an increase in gaze dwell time for ROI of boundary faces. This result indicates that a perceptual discrimination difficulty changed the relative importance (in attentional terms) of facial features. Evidence relates increase in cognitive processing load to difficulty in categorizing boundary faces (Cheetham et al., 2011; Weis & Wiese, 2017; Wiese et al., 2019; Wiese & Weis, 2020). In contrast, whereas in our investigation an increase in dwell time for boundary faces' nose was found, the study of Cheetham et al. (2013) found greater dwell times in boundary faces' eyes and mouth.

This output may occur due to differences in the experimental design. Cheetham et al. (2013) displayed a conventional central fixation point, which was followed by a centralized facial stimulus. Although we presented a centralized stimulus, the fixation point was placed horizontally centered but vertically lateralized (and randomly

[2] Two factors related to the experimental design might explain the low dwell time found for the mouth area in the present study. First, we calculated the total dwell time considering the entire screen and not just the face area. Second, the fixation point preceding the facial stimulus was not centralized, but lateralized, and therefore the fixation point and the stimulus had no overlapping areas (see Discussion section). In addition, all faces are non-expressive and the mouth is a diagnostic feature for emotional processing (Smith et al., 2005).

presented in each hemifield) to avoid bias toward any ROI, especially the nose once their areas overlap. Thus, despite the static nature of the visual stimuli, the participant had to perform an eye movement every time a new face was presented. Võ et al. (2012) investigated gaze allocation in dynamic face perception, and they found an increased fixation on the nose region when a face moved quickly. They suggested the nose would serve as a spatial anchor. The authors argue that the facial gaze allocation depends on the demands of the tasks. Considering that gaze allocation is sensitive to strategies to maximize visual processing (Buchan et al., 2007, 2008), the nose might be more relevant for ambiguous stimuli in a human–avatar face categorization task in a more dynamic setting which requires an eye movement to gaze at a face. Advantageously, given the central position of the nose, facial information is balanced in all directions and visual parafovea encompasses both eye and mouth (Võ et al., 2012). Finally, visual processing is optimized in face recognition tasks when looking just below the eye, that is, between the eyes and the nose tip (Peterson & Eckstein, 2012).

Despite the increased dwell time for the nose of boundary faces, no differences were found regarding stimuli gender, participant gender and its interaction in avatar, boundary, and human faces. Such results do not support our initial predictions. To the best of our knowledge, this is the first study that controlled the face gender to investigate the hierarchical processing of facial features in the DHL. The absence of gender differences in the ROI dwell times may be related to cultural differences on scanning pattern for faces. The effect of culture on face scanning strategies has been well documented (Blais et al., 2008; Lee et al., 2016). In a recent study, Haensel et al. (2020) observed cultural differences in ROI fixation patterns between British and Japanese participants. In addition, a recent study showed no significant gender differences of fixations in the ROI (Sammaknejad et al., 2017). Moreover, while the present study carried out a human–avatar categorization task, studies that investigated gender differences in face perception generally implement gender-related tasks (e.g., male–female categorization). Such differences between tasks have been shown to influence eye movements (Bowers et al., 2021).

Differently from results of dwell time, gender differences were found in the slope. Men showed a sharper slope than women. Such difference could be understood as a representation of categorization strength. Therefore, the higher the slope, the thinner the range in the DHL for the uncanny valley occurrence; this seems to be the case for perceptual categorization in men. In contrast, a lower slope is associated with a smoother transition between categories, and a wider range in the DHL for the uncanny valley occurrence; this seems to be the case for perceptual categorization in women. Apart from the gender difference in the slope, men and women presented sharp S-shaped psychometric curves, which indicates a cognitive process of perceptual categorization.

The present study also showed differences in the RT among experimental conditions of the DHL factor. Our results showed that boundary faces had longer RTs than avatar and human faces, regardless of the participant gender or the face gender. Previous studies found higher RTs for ambiguous categories compared to human and avatar categories (e.g., de Borst & de Gelder, 2015). As stated previously, long latencies relate to complex perceptual tasks, and a high RT surrounding the categorization threshold supports this region in the DHL as the area of the greatest categorization ambiguity (Cheetham et al., 2011, 2013). The bell-shaped curve for RT peaking on the categorization boundary morphing levels may reflect conflict in decision making during a forced-choice categorization task (Cheetham & Jancke, 2013). Studies using the mouse tracking paradigm also support a cognitive conflict around the uncanny valley in the DHL (Weis & Wiese, 2017).

A shorter RT was found for avatar faces in comparison to human and boundary faces. Such a result was observed in previous studies (Cheetham et al., 2011, 2013). Besides categorization ambiguity in boundary faces just mentioned, two hypotheses might explain faster processing of avatar faces. The first hypothesis relates the higher degree of details in human and boundary faces compared to avatar faces. Artificial avatars lack detailed and finer-scaled facial features. In the present study, all texture information was removed from the computational face models to create artificial avatar versions. Thus, it is likely that perception of avatar faces relies more on low spatial frequencies compared to boundary and human faces. According to the fine-to-coarse hypothesis (Hegdé, 2008), which has been consistently supported in face perception studies (e.g., de Moraes et al., 2016; Goffaux et al., 2011), low spatial frequencies conveyed by fast magnocellular pathways are extracted before high spatial frequencies. The precedence to process coarse information may be related to short latencies found when categorizing avatar faces. To elucidate this hypothesis, future investigations must examine the spatial frequency content of the faces and its ROI.

The second hypothesis, the so-called avatar-feature hypothesis by Cheetham et al. (2013), suggests that the decision of categorization in human or avatar is influenced by the strategy that involves establishing the presence or absence of perceptual information that specify an avatar face. Faces would be coded and categorized as "avatar or not avatar" instead of "avatar or human." Supposing that this mechanism of perceptual information involves less cognitive demand, this would provide a categorization advantage in terms of processing time for avatar faces (Cheetham et al., 2013). Although we used the labels "artificial" and "real," instead of "human" and "avatar," respectively, our results are in line with previous results (Cheetham et al., 2011, 2013).

It is worth noting that this study assumed that ambiguous faces would evoke a sense of uncanniness. However, neither stimuli valence nor the feelings of the participants were directly assessed in our experiment and in Cheetham et al. (2013). Nevertheless, many studies support the evidence of a feeling of eeriness associated with the categorization of entities in an uncertainty "gray area" along the DHL (but see MacDorman & Chattopadhyay, 2016). For instance, a study conducted by Burleigh et al. (2013) showed that the DHL is linearly related to emotional response, except for faces located at the categorization boundary, which elicited negative feelings. Yamada et al. (2013) showed that the categorization threshold, the RT peak, and the lowest likability score co-occurred at the same morphing continua range when morphing two of each of real, stuffed, and cartoon human faces. Burleigh and Schoenherr (2015) observed a decrease in participant affinity to a face at a categorization boundary region. Shin et al. (2019) found an increase of eeriness feelings when participants were presented with realistic avatars when compared to cartoon avatars.

Another limitation of the study regards the sample size. It was not possible to collect more data because of the isolation measures taken by the authorities to control the coronavirus disease (COVID-19) pandemic, which included the closure of university facilities and research laboratories. Unfortunately, there are no estimates of reopening at the time we are submitting this paper.

The uncanny valley effect has consistent explanatory hypotheses other than the categorization ambiguity (presented in the Introduction section). To date, besides categorization ambiguity, the perceptual mismatch hypothesis cumulates robust evidence (see Kätsyri et al., 2015; Strait et al., 2017). The perceptual mismatch hypothesis argues that the uncanny valley effect is caused by inconsistencies among specific sensory cues that vary in the DHL. An interesting follow up investigation could frame the perceptual mismatch hypothesis and investigate the hierarchy of facial features along the DHL using eye tracking data. However, it seems to be rather complicated to control stimulus-driven alternative interpretations. A perceptual mismatch is usually implemented as inconsistencies in the realism level of an object's features (e.g., artificial eyes) or due to the presence of atypical features in humans (e.g., large eyes) on an almost human-like face. Therefore, the perceptual salience of artificial or large facial features could drive attention allocation per se.

The present study sought to investigate cognitive aspects of the uncanny valley phenomenon. Relying on the categorization ambiguity theoretical framework, we assumed that higher cognitive demands associated with perceptual ambiguity when categorizing a character as real or artificial could elicit the uncanny valley effect. Here, we investigate whether higher perceptual discrimination difficulty could affect the hierarchical processing of facial features, a robust and very well-known mental operation. Our results showed no stimuli gender influences in the total dwell time for men and women. In addition, no differences were found for the eyes and mouth areas for avatar, boundary, and real face images. However, the results showed greater dwell times for the nose of boundary faces (vs. the nose of avatar and human faces), which is evidence that a perceptual discrimination difficulty changes the allocation of attentional resources and hence alters the visual scan pattern for ambiguous faces in the DHL. Such output is relevant for experimental psychologists and neuroscientists interested in how we interact with and process artificial faces, as well as for graphics developers and designers concerned on how to improve users' experiences from their highly realistic characters, which are increasingly present in everyday life. Our investigation is integrated into a multidisciplinary field called human–computer interaction, which is concerned with the development of realistic human-like characters for movies, games, robots, apps, and chatbots. Future studies on perceptual ambiguity could tackle different visual stimulation (e.g., emotional and dynamic faces) and techniques (e.g., the bubbles technique and the moving window paradigm) to investigate hierarchical processing of facial features in the DHL.

## References

Barton, J. J., Radcliffe, N., Cherkasova, M. V., Edelman, J., & Intriligator, J. M. (2006). Information processing during face recognition: The effects of familiarity, inversion, and morphing on scanning fixations. *Perception*, *35*(8), 1089–1105. https://doi.org/10.1068/p5547

Becker, S. I. (2011). Determinants of dwell time in visual search: Similarity or perceptual difficulty? *PLOS ONE*, *6*(3), Article e17740. https://doi.org/10.1371/journal.pone.0017740

Beesley, T., Pearson, D., & Le Pelley, M. (2019). Eye tracking as a tool for examining cognitive processes. In G. Foster (Ed.), *Biophysical measurement in experimental social science research* (pp. 1–30). Academic Press. https://doi.org/10.1016/B978-0-12-813092-6.00002-2

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLOS ONE*, *3*(8), Article e3022. https://doi.org/10.1371/journal.pone.0003022

Bowers, N. R., Gautier, J., Lin, S., & Roorda, A. (2021). Fixational eye movements depend on task and target. *BioRxiv*. https://doi.org/10.1101/2021.04.14.439841

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Buchan, J. N., Paré, M., & Munhall, K. G. (2007). Spatial statistics of gaze fixations during dynamic face processing. *Social Neuroscience*, *2*(1), 1–13. https://doi.org/10.1080/17470910601043644

Buchan, J. N., Paré, M., & Munhall, K. G. (2008). The effect of varying talker identity and listening conditions on gaze behavior during audiovisual speech perception. *Brain Research*, *1242*, 162–171. https://doi.org/10.1016/j.brainres.2008.06.083

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

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

Cheetham, M., & Jancke, L. (2013). Perceptual and category processing of the uncanny valley hypothesis' dimension of human likeness: Some methodological issues. *Journal of Visualized Experiments*, *4375*(76), Article e4375. https://doi.org/10.3791/4375

Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., & Jancke, L. (2013). Category processing and the human likeness dimension of the uncanny valley hypothesis: Eye-tracking data. *Frontiers in Psychology*, *4*, Article 108. https://doi.org/10.3389/fpsyg.2013.00108

Cheetham, M., Suter, P., & Jäncke, L. (2011). The human likeness dimension of the "uncanny valley hypothesis": Behavioral and functional MRI findings. *Frontiers in Human Neuroscience*, *5*, Article 126. https://doi.org/10.3389/fnhum.2011.00126

Cheetham, M., Suter, P., & Jancke, L. (2014). Perceptual discrimination difficulty and familiarity in the uncanny valley: More like a "Happy Valley." *Frontiers in Psychology*, *5*, Article 1219. https://doi.org/10.3389/fpsyg.2014.01219

Chuk, T., Chan, A. B., & Hsiao, J. H. (2017). Is having similar eye movement patterns during face learning and recognition beneficial for recognition performance? Evidence from hidden Markov modeling. *Vision Research*, *141*, 204–216. https://doi.org/10.1016/j.visres.2017.03.010

Cintra, P. H. P. (2020). *Uncanny valley* [Computer software]. GitHub. https://github.com/PedroHPCintra/Uncanny-Valley

Coutrot, A., Binetti, N., Harrison, C., Mareschal, I., & Johnston, A. (2016). Face exploration dynamics differentiate men and women. *Journal of Vision*, *16*(14), Article 16. https://doi.org/10.1167/16.14.16

de Borst, A. W., & de Gelder, B. (2015). Is it the real deal? Perception of virtual characters versus humans: An affective cognitive neuroscience perspective. *Frontiers in Psychology*, *6*, Article 576. https://doi.org/10.3389/fpsyg.2015.00576

de Moraes, R., Jr., Kauffmann, L., Fukusima, S. S., & Faubert, J. (2016). Behavioral evidence for a predominant and nonlateralized coarse-to-fine encoding for face categorization. *Psychology & Neuroscience*, *9*(4), 399–410. https://doi.org/10.1037/pne0000065

Fabri, M., Moore, D., & Hobbs, D. (2004). Mediating the expression of emotion in educational collaborative virtual environments: An experimental study. *Virtual Reality*, *7*, 66–81. https://doi.org/10.1007/s10055-003-0116-7

Fraser, I. H., & Parker, D. M. (1986). Reaction time measures of feature saliency in a perceptual integration task. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. Young (Eds.), *Aspects of face* *processing* (pp. 45–52). Springer. https://doi.org/10.1007/978-94-009-4420-6_4

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Godard, O., & Fiori, N. (2012). Sex and hemispheric differences in facial invariants extraction. *Laterality: Asymmetries of Body, Brain, and Cognition*, *17*(2), 202–216. https://doi.org/10.1080/1357650X.2011.556641

Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2011). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebral Cortex*, *21*(2), 467–476. https://doi.org/10.1093/cercor/bhq112

Goodman, L. R., Phelan, H. L., & Johnson, S. A. (2012). Sex differences for the recognition of direct versus averted gaze faces. *Memory*, *20*(3), 199–209. https://doi.org/10.1080/09658211.2011.651089

Haensel, J. X., Ishikawa, M., Itakura, S., Smith, T. J., & Senju, A. (2020). Cultural influences on face scanning are consistent across infancy and adulthood. *Infant Behavior and Development*, *61*, Article 101503. https://doi.org/10.1016/j.infbeh.2020.101503

Hall, J. K., Hutton, S. B., & Morgan, M. J. (2010). Sex differences in scanning faces: Does attention to the eyes explain female superiority in facial expression recognition? *Cognition and Emotion*, *24*(4), 629–637. https://doi.org/10.1080/02699930902906882

Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience*, *9*(6), 467–479. https://doi.org/10.1038/nrn2374

Hegdé, J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in Neurobiology*, *84*(4), 405–439. https://doi.org/10.1016/j.pneurobio.2007.09.001

Heisz, J. J., Pottruff, M. M., & Shore, D. I. (2013). Females scan more than males: A potential mechanism for sex differences in recognition memory. *Psychological Science*, *24*(7), 1157–1163. https://doi.org/10.1177/0956797612468281

Herlitz, A., & Lovén, J. (2013). Sex differences and the own-gender bias in face recognition: A meta-analytic review. *Visual Cognition*, *21*(9–10), 1306–1336. https://doi.org/10.1080/13506285.2013.823140

Hessels, R. S., Niehorster, D. C., Nyström, M., Andersson, R., & Hooge, I. T. C. (2018). Is the eye-movement field confused about fixations and saccades? A survey among 124 researchers. *Royal Society Open Science*, *5*(8), Article 180502. https://doi.org/10.1098/rsos.180502

Ho, C.-C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, *26*(6), 1508–1518. https://doi.org/10.1016/j.chb.2010.05.015

Ho, C.-C., & MacDorman, K. F. (2017). Measuring the uncanny valley effect: Refinements to indices for perceived humanness, attractiveness, and eeriness. *International Journal of Social Robotics*, *9*(1), 129–139. https://doi.org/10.1007/s12369-016-0380-9

The jamovi Project. (2020). *jamovi* (Version 1.6) [Computer software]. https://www.jamovi.org

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. https://doi.org/10.1037/0033-295X.87.4.329

Kätsyri, J., Förger, K., Mäkäräinen, M., & 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. *Frontiers in Psychology*, *6*, Article 390. https://doi.org/10.3389/fpsyg.2015.00390

Kleiner, M., Brainard, D., & Pelli, D. (2007). *What's new in psychtoolbox-3?* [Conference session]. Perception Supplement—Thirtieth European Conference on Visual Perception, Arezzo, Italy. https://doi.org/10.1177/03010066070360S101

Lee, Y. J., Greene, H. H., Tsai, C. W., & Chou, Y. J. (2016). Differences in sequential eye movement behavior between Taiwanese and American viewers. *Frontiers in Psychology*, *7*, Article 697. https://doi.org/10.3389/fpsyg.2016.00697

Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition—Women's faces make the difference. *Brain and Cognition*, *50*(1), 121–128. https://doi.org/10.1016/S0278-2626(02)00016-7

Lovén, J., Herlitz, A., & Rehnman, J. (2011). Women's own-gender bias in face recognition memory. *Experimental Psychology*, *58*(4), 333–340. https://doi.org/10.1027/1618-3169/a000100

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska directed emotional faces*. Karolinska Institutet, Department of Clinical Neuroscience, Psychology Section.

MacDorman, K. F., & Chattopadhyay, D. (2016). Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition*, *146*, 190–205. https://doi.org/10.1016/j.cognition.2015.09.019

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

MacDorman, K. F., Vasudevan, S. K., & Ho, C.-C. (2009). Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *AI & Society*, *23*(4), 485–510. https://doi.org/10.1007/s00146-008-0181-2

Megreya, A. M., Bindemann, M., & Havard, C. (2011). Sex differences in unfamiliar face identification: Evidence from matching tasks. *Acta Psychologica*, *137*(1), 83–89. https://doi.org/10.1016/j.actpsy.2011.03.003

Moore, R. K. (2012). A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena. *Scientific Reports*, *2*, Article 864. https://doi.org/10.1038/srep00864

Mori, M. (2012). The uncanny valley In K. F. MacDorman, & N. Kageki (Trans.), *IEEE Robotics and Automation* (Vol. 19, No. 2, pp. 98–100). IEEE. (Original work published 1970). https://doi.org/10.1109/MRA.2012.2192811

Orquin, J. L., & Holmqvist, K. (2018). Threats to the validity of eye-movement research in psychology. *Behavior Research Methods*, *50*(4), 1645–1656. https://doi.org/10.3758/s13428-017-0998-z

Pavlova, M. A., Mayer, A., Hösl, F., & Sokolov, A. N. (2016). Faces on her and his mind: Female and likable. *PLOS ONE*, *11*(6), Article e0157636. https://doi.org/10.1371/journal.pone.0157636

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(48), E3314–E3323. https://doi.org/10.1073/pnas.1214269109

Pollick, F. E. (2010). In search of the uncanny valley. In P. Daras, & O. M. Ibarra (Eds.), *User centric media. UCMEDIA 2009. Lecture notes of the institute for computer sciences, social informatics and telecommunications engineering* (Vol. 40, pp. 69–78). Springer. https://doi.org/10.1007/978-3-642-12630-7_8

Rehnman, J., & Herlitz, A. (2006). Higher face recognition ability in girls: Magnified by own-sex and own-ethnicity bias. *Memory*, *14*(3), 289–296. https://doi.org/10.1080/09658210500233581

Sammaknejad, N., Pouretemad, H., Eslahchi, C., Salahirad, A., & Alinejad, A. (2017). Gender classification based on eye movements: A processing effect during passive face viewing. *Advances in Cognitive Psychology*, *13*(3), 232–240. https://doi.org/10.5709/acp-0223-1

Shen, J., Reingold, E. M., Pomplun, M., & Williams, D. E. (2003). Saccadic selectivity during visual search: The influence of central processing difficulty. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye* (pp. 65–88). North-Holland Publishing. https://doi.org/10.1016/B978-044451020-4/50005-0

Shin, M., Kim, S. J., & Biocca, F. (2019). The uncanny valley: No need for any further judgments when an avatar looks eerie. *Computers in Human Behavior*, *94*, 100–109. https://doi.org/10.1016/j.chb.2019.01.016

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science*, *16*(3), 184–189. https://doi.org/10.1111/j.0956-7976.2005.00801.x

Sommer, W., Hildebrandt, A., Kunina-Habenicht, O., Schacht, A., & Wilhelm, O. (2013). Sex differences in face cognition. *Acta Psychologica*, *142*(1), 62–73. https://doi.org/10.1016/j.actpsy.2012.11.001

Strait, M. K., Floerke, V. A., Ju, W., Maddox, K., Remedios, J. D., Jung, M. F., & Urry, H. L. (2017). Understanding the uncanny: Both atypical features and category ambiguity provoke aversion toward humanlike robots. *Frontiers in Psychology*, *8*, Article 1366. https://doi.org/10.3389/fpsyg.2017.01366

Turner, B. M., & Van Zandt, T. (2012). A tutorial on approximate Bayesian computation. *Journal of Mathematical Psychology*, *56*(2), 69–85. https://doi.org/10.1016/j.jmp.2012.02.005

Urgen, B. A., Kutas, M., & Saygin, A. P. (2018). Uncanny valley as a window into predictive processing in the social brain. *Neuropsychologia*, *114*, 181–185. https://doi.org/10.1016/j.neuropsychologia.2018.04.027

Võ, M. L. H., Smith, T. J., Mital, P. K., & Henderson, J. M. (2012). Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *Journal of Vision*, *12*(13), Article 3. https://doi.org/10.1167/12.13.3

Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology*, *19*(4), 393–407. https://doi.org/10.1037/gpr0000056

Weis, P. P., & Wiese, E. (2017). Cognitive conflict as possible origin of the uncanny valley. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 1599–1603. https://doi.org/10.1177/1541931213601763

Wiese, E., Mandell, A., Shaw, T., & Smith, M. (2019). Implicit mind perception alters vigilance performance because of cognitive conflict processing. *Journal of Experimental Psychology: Applied*, *25*(1), 25–40. https://doi.org/10.1037/xap0000186

Wiese, E., & Weis, P. P. (2020). It matters to me if you are human: Examining categorical perception in human and nonhuman agents. *International Journal of Human-Computer Studies*, *133*, 1–12. https://doi.org/10.1016/j.ijhcs.2019.08.002

Wolff, N., Kemter, K., Schweinberger, S. R., & Wiese, H. (2014). What drives social in-group biases in face recognition memory? ERP evidence from the own-gender bias. *Social Cognitive and Affective Neuroscience*, *9*(5), 580–590. https://doi.org/10.1093/scan/nst024

Wright, D. B., & Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. *Acta Psychologica*, *114*(1), 101–114. https://doi.org/10.1016/S0001-6918(03)00052-0

Yamada, Y., Kawabe, T., & Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *The Japanese Psychological Research*, *55*(1), 20–32. https://doi.org/10.1111/j.1468-5884.2012.00538.x

Received May 4, 2021
Revision received October 22, 2021
Accepted November 29, 2021 ■