

Investigating the Uncanny Valley Phenomenon Through the Temporal Dynamics of Neural Responses to Virtual Characters

Chiara Gorlini
Politecnico di Milano
Milan, Italy
chiara.gorlini@mail.polimi.it

Laurits Diken
IT-University Of Copenhagen
Copenhagen, Denmark
ldix@itu.dk

Paolo Burelli
IT-University Of Copenhagen
Copenhagen, Denmark
pabu@itu.dk

studying uncanny valley for the sake of game design, this is why they use fictional characters;
are we justified in using fictional characters?

Abstract—The Uncanny Valley phenomenon refers to the feeling of unease that arises when interacting with characters that appear almost, but not quite, human-like. First theorised by Masahiro Mori in 1970, it has since been widely observed in different contexts from humanoid robots to video games, in which it can result in players feeling uncomfortable or disconnected from the game, leading to a lack of immersion and potentially reducing the overall enjoyment. The phenomenon has been observed and described mostly through behavioural studies based on self-reported scales of uncanny feeling: however, there is still no consensus on its cognitive and perceptual origins, which limits our understanding of its impact on player experience. In this paper, we present a study aimed at identifying the mechanisms that trigger the uncanny response by collecting and analysing both self-reported feedback and EEG data.

Index Terms—Uncanny Valley, Virtual Characters, Video Games, EEG, ERP, Survey, Perception, Cognition

I. INTRODUCTION

The concept of the Uncanny Valley (UV) refers to the phenomenon by which an artificial or virtual human-like entity is increasingly unsettling as it becomes more realistic but does not quite achieve a perfect human likeness. This effect can be observed in many media, including games, and has important implications for game design and player experience [29].

Studying this phenomenon can help game designers and developers to create more realistic and believable characters without crossing into the UV, which can lead to negative player reactions and reduced immersion. By understanding the factors that contribute to the effect, teams can make informed decisions about how to design their characters and game worlds.

Moreover, by studying the phenomenon, we can better understand the psychological processes that underlie our perception of virtual characters. Research has shown that the effect is linked to our innate ability to recognize human facial expressions and interpret emotional cues [30]. By studying the UV in games, we can gain insight into how our brains process visual information and how we form emotional connections with virtual characters.

The most common approach to studying the empirical basis of this phenomenon relies on self-reported subjective

measurements, in which a series of stimuli is followed by questionnaires. These studies have shown repeatedly that some form of eeriness or unease appears to be triggered by different types of semi-realistic stimuli; however, the self-reported nature of these studies comes with some major limitations [34].

Self-reported feedback is subjective and can be influenced by a variety of factors, such as individual preferences, biases, and cultural differences. Additionally, self-reported feedback may not always accurately reflect a person's actual emotional response, as people may not always be aware of or able to articulate their feelings in a precise way.

Moreover, the lack of uniformity of the stimuli and questionnaires used makes it difficult to compare results across studies. The questionnaires used to measure the UV effect do not always attempt to measure the same concept and define the uncanny feeling in very different ways. This can lead to inconsistencies in the results and make it challenging to draw definitive conclusions about the effect. Therefore, while self-reported studies can provide valuable insights into the UV effect, they should be supplemented with other objective measures, such as physiological responses or behavioural observations to obtain a more comprehensive understanding of how people react to artificial stimuli and understand its cognitive origins [5].

In the last decade, in an attempt to address these limitations, researchers have started to investigate how brain activity is related to the UV. These studies have shown the potential to shed light on the causes of the effect based on more objective measurements such as EEG (electroencephalography) or fMRI (functional magnetic resonance imaging). Both are non-invasive methods to record, directly or indirectly, electrical activity in the brain allowing researchers to gain insight into how the brain processes and responds to stimuli [32].

These studies have investigated both static and animated characters and have analysed their effect on the participants' brains primarily against different Event-Related Potentials (ERPs), with a few studies crossing brain activity measurements with behavioural variables [1], [4].

Based on these results, in this paper, we present a two-steps study in which we first build a validated stimuli dataset based

on self-reported subjective measurements collected through an online survey and then investigate the potential origins of these subjective measurements in a laboratory experiment in which we collect EEG responses to the validated stimuli. By collecting both self-reported and psychophysiology measurements and by analysing the dynamics of the EEG response, we aim at finding objective evidence that can connect the two types of responses and that can either confirm or confute the different hypotheses on the origins of the UV phenomenon.

II. BACKGROUND

Mashiro Mori theorised the Uncanny Valley phenomenon in 1970 to describe the unsettling feeling humans experience when looking at robots resembling human characteristics too closely [19]. The theory states that there is a non-linear relationship between the realism of an artificial character and its likability characterised by an evident dip at a level of realism that is almost but not quite perfect. While initially observed with humanoid robots, the rise of computer-generated images in video games and movies has opened a new area of application of the phenomenon, sparking new research aimed at finding behavioural and cognitive evidence of the UV [30].

Over the years, different theories have formed to explain the UV phenomenon that can be roughly separated into two typologies. Firstly, the ones that explain the uncanny feeling as an instinctual response to stimuli early in brain perception (i.e. Pathogen Avoidance, Mortality Salience and Evolutionary Aesthetics) [34]. The second class of theories explains the UV as a result of the cognitive process that occurs later in brain perception (i.e. Categorical Uncertainty, Violation of Expectation, Mind Perception, Dehumanization) [5], [20], [31]. These latter cognitive process hypotheses have received more credits in the literature, but there is still no clear evidence to fully support either of them.

The primary mean of studying the UV is through self-reported subjective questionnaires. In these studies on the phenomenon, researchers manipulate a character's level of human photo-realism and measure the character's perceived humanness. Behavioural studies on UV mainly propose faces (still images and videos) as the stimuli to investigate. However, experiments differ in stimulus creation techniques for human-likeness modulations in robots, virtual agents, avatars, and perceptual scale modalities [8].

Stimulus creation techniques are strictly controlled, ranging from artificial to entirely realistic characters (i.e. face distortion, realism render, and morphing) [16], [18], [24], or distinct entities based on the selection of existing computer-animated characters or robots [21].

Different perceptual exploratory constructs are used to describe the UV, such as eeriness, warmth, perceived threat, likability, and familiarity, whose relation can give insight into the phenomenon [8]. The dimensions are measured through a variety of terms used to define semantic differential scales. For example, MacDorman et al. [17] define the eeriness dimension as composed of 9 scales, each represented between two descriptors, Shin et al. [26] consider the same dimension but use

four descriptive scales, and Rosenthal-von der Pütten [21] do not consider eeriness but perceived threat as a main dimension.

The variability between studies in terms of stimuli, scales and observations leads to the formulation of different results and causal explanations of the phenomenon. The UV shape varies between studies identifying linear relationships between human-likeness and affinity, 'weak uncanny valley', and 'strong uncanny valley' where the lowest affinity is predicted for high levels of human-likeness [15]. The different findings suggest that the UV is a multidimensional construct influenced by various factors, including stimulus creation techniques, non-trivial questionnaire terms, different levels of familiarity for characters, and other facets that might make entities unpleasant [8].

Although subjective questionnaires are a valuable tool for investigating the perception of characters, the complexity of the phenomenon makes them insufficient to explain the underlying causes of the UV. The human brain processes the faces between the onset of the visual stimulation and the behavioural response for several hundreds of milliseconds. Clarifying the exact time course of face processing by monitoring brain activity may give more insight into the reasoning behind the UV phenomenon.

There are few studies on brain activity related to UV in literature so far, see [32] for a recent comprehensive review. Electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) have been used to find valuable biological markers and brain regions that might be involved in the face and object processing and the UV phenomenon.

Three EEG studies have been published using still images as stimuli. These studies use event-related potentials (LPP, N170, P200 and P300) as their main dependent variable for the different categories of stimuli [5], [25]. ERPs are peaks in the electrical activity of the brain that reflect different stages of information processing and are usually identified by their polarity and time of occurrence. Dynamic stimuli have also been used to test the hypothesised exacerbation of the UV curve. In EEG studies, the ERP adopted for such investigation is the N400 [31].

One under-investigated aspect in the current literature is the comparison of early vs late processing [32], [34] and at what time the uncanny response occurs. This is an important question since the temporal dimension might help to discern the course of the feeling as cognitive mechanisms happen at different time points in the perceptual process.

In general, theories can be divided into perceptual processing theories (e.g. pathogen avoidance or mortality salience) and cognitive processing theories (e.g. violation of expectations or category uncertainty) [34]. If the UV phenomenon relied on purely perceptual differences, we would expect to see the early ERP components being modulated when the subject is experiencing uncanniness instead of the later ones.

The present study focuses on two major ERP components: an early N170 component and a late N400 component. The N170 is specifically associated with processing human faces [23]. It is characterised by a low negative peak at

the temporal lobe around 170 ms post-stimulus onset. The N170 is actually a special case of the visual ERP component called N1 but with a larger amplitude than non-face objects. Researchers are still discussing the neurological underpinnings of this 'N170 face effect', but there is a general agreement that the increased amplitude is evidence of more synchronised processing in the temporal lobe for facial stimuli [13].

In the context of UV, the N170 peak seems sensitive to facial human likeness and has a role in face categorisation. In particular, Schindler et al. [25] find a larger N170 amplitude for cartoon and natural human faces than for avatar faces. It is still uncertain how this early visual component is related to the UV phenomenon. The other ERP component investigated in this study is the N400. This component has been widely used to identify semantic incongruities in the language – e.g. “I like my coffee with cream and sugar/dog” [14]. The N400 is characterised by negative activity around the centro-parietal areas between 200 and 600 ms post-stimulation, particularly around 400 ms.

It has been found in previous studies that increased N400 amplitude is related to robot-like movements of human-looking androids [31] and talking CG characters [20]. The common interpretation is that the artificial movements trigger an error correction response like a word in the wrong semantic context. It remains unknown if still images would elicit the same response, and be a general electrophysiological marker of the UV hypothesis or if it simply stems from the unexpectedness of artificial motions.

The present study is inspired by the work of Wang et al. [33], in which they explored the reliability of the Dehumanisation Hypothesis through behavioural analysis. The Dehumanization hypothesis explains the uncanny feeling through the initial tendency to over-attribute a mind to nonhuman agents and then the perception of the anthropomorphise human replica as lacking humanness.

Since their promising results demonstrate that the process of dehumanisation is more likely to account for the perceived uncanniness of android faces, the present study investigates the same hypothesis supported by an EEG study. By analysing the temporal dynamics of face processing through the study of both early and late ERPs, we aim at testing the Dehumanisation Hypothesis and uncovering the underlying mechanisms of the Uncanny Valley phenomenon.

III. METHODS

The study is structured in two phases. In the first phase, we designed a questionnaire to collect self-reported responses to different kinds of virtual characters. This questionnaire aims at building a baseline of the different dimensions describing the “uncanny” feeling. We use the responses to fine-tune and validate the images that are used for the second phase of the study.

In the study’s second phase, we conduct a laboratory experiment, in which we expose a small sample of participants to a selection of previously validated virtual characters and analyse the dynamics of their EEG response in terms of

event-related potentials (ERPs). By analyzing the ERPs of the participants in response to the virtual characters, we aim to identify which sequences of neural activations the different kinds of images trigger in the brain and, based on this, identify specific cognitive processes that are connected to the UV.

A. Questionnaire

The questionnaire¹ is designed to collect reported feelings about a series of images. Every participant is first welcomed by a brief explanation of the study. After this first step, they are presented with a page that briefs them with some definitions and, after having given their consent, from this page, they are guided through a randomised sample of 20 images. For each image, their response is collected as a series of 7-point Likert scales. In the end, each participant is shown a debriefing page.

1) *Stimulus set*: Each participant is presented with a sequence of 20 images to rate, these images are a randomised sample from the full image dataset used in the questionnaire. The full dataset is composed of 109 images divided into three categories: *unrealistic*, *semi-realistic* and *realistic*. The categorisation of the images is a direct consequence of the non-linear emotional response to realism that defines the UV phenomenon.

The first category (Figure 1a) includes unrealistic characters with exaggerated features often seen in the style of children’s animated movies. These characters are rarely thought of as uncanny, despite clearly acting humanly. We picked the 36 characters belonging to this category from successful animation films and video games. The second category (Figure 1b) includes animated synthetic characters with proportional human features designed to look realistic but don’t achieve full realism. These are the characters which are more often described as uncanny, eliciting a negative emotional reaction regardless of the role they play. We picked 37 characters from video games and movies that have been described as uncanny on online communities and websites discussing the phenomenon – e.g. <http://vtvtropes.org>. The last category (Figure 1c) includes a set of 36 face images of real people front-facing the camera. The images are selected from the Face Research Lab London Set [6] with various gender and age combinations.

Inclusion criteria for the images are: high picture quality cropped and zoomed on the face of the character, and the character must show at least 3/4 of the face and have a neutral expression. The *Unrealistic* and *Semi-realistic* categories include characters from well-known digital productions, but rarely the main characters to avoid strong familiarity influencing the study.

The images are processed to have the same resolution and size, their contrast and luminance are balanced, and each face is cut and pasted on a grey background. This is done to minimize low-level visual stimulus differences such as

¹The source code of the questionnaire, the data and the statistical analysis are available at <https://github.com/itubrainlab/uncanny-face-questionnaire/tree/COG-2023>

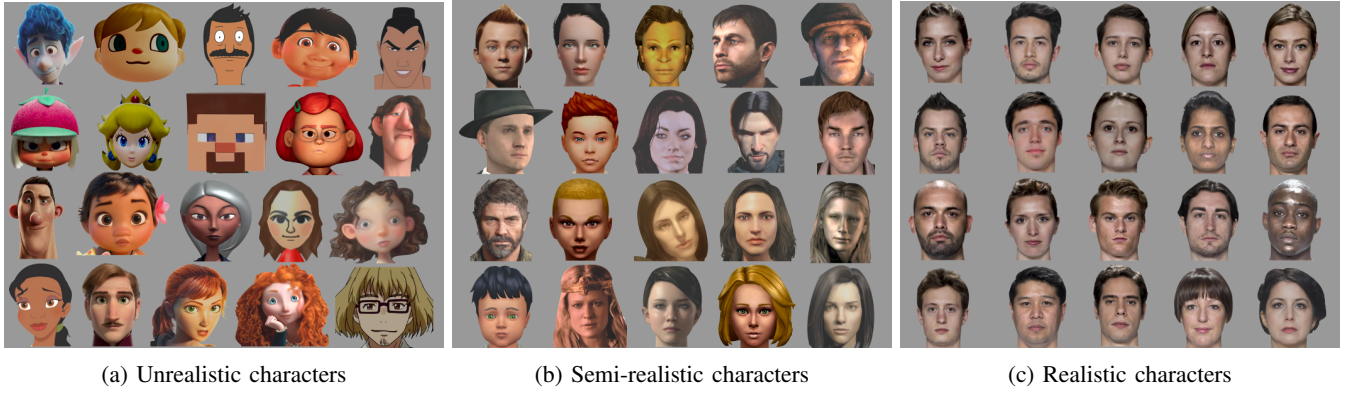


Fig. 1: Images used as stimuli in the EEG study. These images have been selected through a prior online questionnaire and they are organised into three categories: *unrealistic*, *semi-realistic* and *realistic* characters.

luminance and contrast to mitigate these differences between stimuli in the EEG part of the experiment.

2) *Structure, questions and hypotheses*: According to the UV hypothesis [34], our three categories should fall on three different spots on the likability vs. human-likeness scale. The images in the *Unrealistic* category should have low human likeness and medium to high likability. The human category should be high in both likability and human likeness. Lastly, the semi-realistic category should have high human likeness but low likability. This is the hypothesised outcome of the questionnaire.

To operationalise the two dimensions of the UV effect, we construct three multi-item scales inspired by a recent literature review of Uncanny Valley [8] studies. Each item consists of an anchor word at each end and a short description of the word. The participants are asked to rate the character on a seven-point Likert scale with the two anchor words as the outer points for each item.

The structure of the scales and the items is the following:

- **Realism** scale: *Fictional/Real* [-3, +3], and *Human-made/Human-like* [-3,+3];
- **Eeriness** scale: *Ordinary/Eerie* [0,+6], *Plain/Unsettling* [0,+6], *Dull/Creepy* [0,+6], and *Unemotional/Hair-raising* [0,+6];
- **Warmth** scale: *Hostile/Friendly* [-3,+3], *Grumpy/Cheerful* [-3,+3] and *Cold-hearted/Warm-hearted* [-3,+3].

Seven of the nine items are meant to ascertain the “uncanniness” of the character; these items are divided into a *Warmth* scale and an *Eeriness* scale, largely inspired by previous work [8], [10]. The *Eeriness* scale is running from a neutral word to a negative word, intended to capture the negative aspect of uncanny feeling. The selection of terms aims at preserving the hypothesized construct of eeriness, which encompasses both fright and spine-tingling feelings. The specific terms are adapted from Ho and MacDorman [10], with a few adaptations to apply it to a static image dataset.

The *Warmth* scale is symmetrical around zero. Diel et al. [8] find that this scale is not directly connected to the UV effect;

however, we include it in the study as it has the potential to be an effective measure to discriminate the *unreal* category [8]. To measure the independent variable *Realism*, we adapt the term to better suit animated characters specifically resulting in two items describing the realism of the character and how human-like it seems.

3) *Analysis*: Since the questionnaire is composed of three multi-item psychometric scales, the first step in the analysis consists in verifying the structure of these scales through confirmatory factor analysis [3]. First, to verify the existence of multi-item factors, we conduct a Bartlett’s test of Sphericity; second, we calculate the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy to examine the strength of the partial correlations between the items. Finally, we fit a Confirmatory Factor Analysis model to estimate the loadings of the items with respect to the scales.

For each scale verified in the aforementioned analysis, we calculate their means across categories and verify the presence of significant differences through a series of non-parametric Mann-Whitney U tests. Based on the results of this analysis, we select a subset of 20 images for each category that perform better in terms of “uncanny” feeling to be used for the EEG Experiment. The reason for this sampling is to make sure that only images that are validated to follow the UV curve are used as stimuli in the experiment; this way, we can increase the likelihood to capture the phenomenon in the participants’ psychophysiological responses.

B. EEG Experiment

The purpose of this experiment is to collect the EEG response to the images selected through the survey. The testing procedure² to collect this data is the following: first, the participants are welcomed into the testing room and received a short briefing; second, they are seated in front of a screen and prepared for the recording. After the preparation, each participant goes through two 20 minutes of recording sessions separated by a short break to prevent fatigue.

²The source code of the EEG experiment and the analysis are available at <https://github.com/itubrainlab/uncanny-face-eeeg/tree/COG-2023>

Each session contains 240 trials in randomised order. With 20 images in each stimulus condition, each participant sees every image four times. The participants are instructed to restrain from **blinking** during the stimulus time, and breaks of 10 seconds every 20 trials are added to minimise eye strain.

During each trial, the participant is first exposed to a centered fixation cross for 500 ms followed by a period of stimulus onset asynchrony (SOA) randomised between 300 and 600 ms. Afterwards, the stimulus is shown for 700 ms. The participant is then asked to rate the character just seen in terms of perceived animacy on a 1-3 scale. The participant uses their left hand and has a maximum of 2000 ms to rate.

The purpose of the task is to keep the participant alert; we chose 'animacy' as it relates directly to the dehumanisation hypothesis [33], [34], this way the participant hopefully was constantly thinking about the animacy of the character, which should drive the dehumanisation process. Furthermore, the task was chosen to be as simple as possible to not disturb the EEG signal with motor or pre-motor activity in preparation for the reporting movement. The analysis of this behavioural task is not of major importance in this article's analysis, since we have poor granularity with only three options.

1) *Equipment*: The EEG recording was executed using a g.tec³ 32 channel g.USBAMP amplifier, with g.LADYbird active electrodes arranged in accordance with the 10/20 system [12] and referenced to an earlobe electrode. The placement of the electrodes can be seen in Figure 2. Data was recorded using MathWorks Simulink⁴ with a 256 Hz sampling frequency filtered with a 0.1 Hz high-pass filter and a 50 Hz Notch filter to remove drift and power line noise.

2) *Pre-processing*: After the recording, the EEG data is further filtered using a FIR band-pass filter at 1 Hz to 35 Hz. Then flat and noisy channels are detected and removed from further analysis with the RANSAC algorithm, following the procedure of the PREP EEG pre-processing pipeline [2]. The eye blink artefacts are removed through Independent Component Analysis (ICA) [28], training the ICA on four-second intervals [7]. The data was split into epochs from -200 ms before to 700 ms past the stimulus onset and baselined on the pre-stimulus part of the epoch. Again following the PREP pipeline, the Autoreject [11] algorithm is used to detect and interpolate or reject bad epochs. EEG pre-processing and analysis are based on the MNE python package [9].

3) *Small N studies and within-participant design*: The EEG analysis in this study features a relatively small number of participants, which would normally limit the generalisability of the conclusions drawn from the data. However, relying on within-participant analysis rather than group-level analysis can increase the robustness of the conclusions significantly [27]. Using the singular participant as the replicational unit decreases the risk of type I errors since we rely on the same pattern showing up in each participant independently.

4) *ERP analysis*: Two ERP components analysed in this study are the N170 and N400. An upside of small N studies

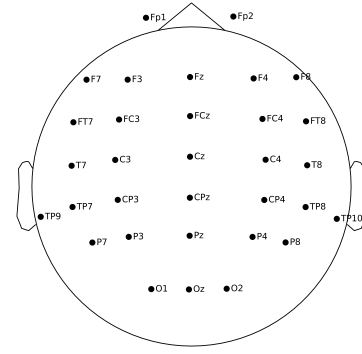


Fig. 2: EEG montage displaying the placement of the 32 electrodes.

is that they simplify the customisation of the analysis to individual differences between participants; in this study, for instance, we can personalise the time windows of analysis for the N170 components of each participant to accommodate the individual variability of the onset and peak of the component. To prevent leakage into our statistical analysis, the time windows are picked based on the evoked potential averaged on all conditions.

Contrarily to the N170, the N400 component is expected to be present only in one condition; therefore, we cannot pick time windows based on the average signal and we cannot have individualised time windows. Instead, the time window was set at 450-600 ms for all participants. To measure the N170, we use the singular electrode P8, which has been shown to consistently show the N170 face effect, re-referencing the signal to the average amplitude [23]. To measure the N400, we use the centro-parietal electrodes: CPz, CP3, CPz, CP4, P3, Pz, P4 [14].

We are using the *realistic* human faces as the baseline condition, so the hypotheses are stated as a difference to this condition. We are testing two main hypotheses:

- H1 There is a significantly stronger N170 response to *un-realistic* characters but not the *semi-realistic* ones when compared to *realistic* characters.
- H2 There is a significantly stronger N400 response to *semi-realistic* characters but not the *unrealistic* ones when compared to *realistic* characters.

IV. RESULTS

Following the structure of the study, this section is divided into two parts. In the first one, we present the results collected and analysed from the questionnaire and we describe the dataset selected for the EEG experiment. In the second part, we present the result and the analysis of the EEG data collected in the experiment.

Through the online questionnaire, we collected responses from 273 participants. The participants' age is between 17 and 57 years with a mean of approximately 29 with around 56%

³<https://www.gtec.at/product/gusbamp-research/>

⁴<https://www.mathworks.com/products/simulink.html>

	Realism	Uncanny	Warmth
Fictional/Real	2.60	0.00	0.00
Human-made/Human-like	2.06	0.00	0.00
Eerie/Ordinary	0.00	1.56	0.00
Unsettling/Plain	0.00	1.68	0.00
Creepy/Dull	0.00	1.57	0.00
Hair-raising/Unemotional	0.00	1.21	0.00
Hostile/Friendly	0.00	0.00	1.37
Grumpy/Cheerful	0.00	0.00	1.21
Cold-hearted/Warm-hearted	0.00	0.00	1.31

TABLE I: Confirmatory factor analysis loadings of the scales and their items in the questionnaire.

of the sample being composed of non-native English speakers. Each image has been rated between 20 and 50 times.

In the confirmatory factor analysis, the Bartlett's test of Sphericity results in a chi-square value of 18964.41 p-value equal to 0.0. This indicates that the dataset has a lower dimensionality compared to the number of items, a necessary condition of the existence of the multi-item scales. The result of the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is 0.815, which indicates the presence of strong partial correlations between the items. Finally, the confirmatory factor analysis shown in Table I, confirms the presence of three independent factors: *Realism*, *Eeriness* and *Warmth*, corresponding to the three multi-items scales designed in the questionnaire. For each factor, we calculate a score equal to the sum of the items' values, with a range between -6 and $+6$ for the *Realism* scale, 0 and 24 for the *Eeriness* scale, and -9 and $+9$ for the *Warmth* scale.

Based on the Eeriness scale, a subset of 20 images from each category is selected. This subset contains images that maximise the difference in reported eeriness feeling between the *semi-realistic* category and the other two. Through this criterion, we aim at selecting images with a stronger potential to generate an uncanny response that follows the canonical Uncanny Valley.

A non-parametric Mann-Whitney U test was conducted on these images to verify whether the categories reflected their respective labels. Table II shows the results obtained.

	p-value	Median Score	Score Scale
<i>Eeriness</i>	< 0.01	7, 12, 3	[0,+24]
<i>Warmth</i>	< 0.01	4, -1, 1	[-9,+9]
<i>Realism</i>	< 0.01	-6, -4, 6	[-6,+6]

TABLE II: Non-parametric Mann-Whitney U test results. The Score Scale is computed by summing the items' values in the specific scale. The Median Scores are reported for *Unrealistic*, *Semi-realistic* and *Realistic* categories, respectively.

A. ERP analysis

Five people (1 female) participated in the EEG experiment. They were all sampled from the staff at the IT University of Copenhagen. Crucially, they were blind to the experiment's purpose and hypothesis and had also not participated in the

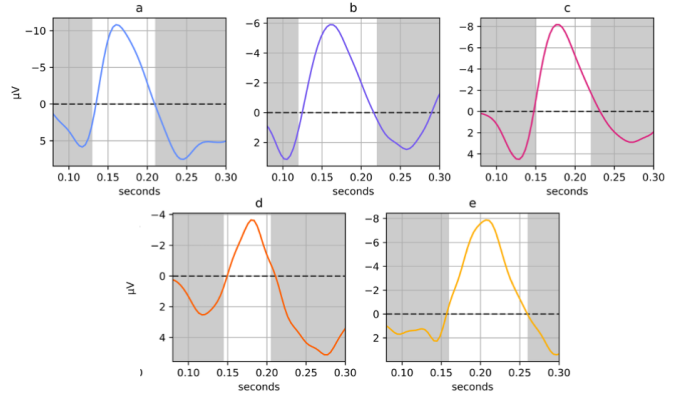


Fig. 3: N170 time windows estimated for the five participants in the EEG experiment.

questionnaire described above. The participants are labelled 'a', 'b', 'c', 'd' and 'e' in the remainder of the text⁵.

For each participant, we analyse the characteristics of their cognitive response based on the intensity of their N170 and N400 ERPs. As described in Section III-B4, the time windows for the N170 component are personalised for each participant based on their average ERP across all experimental conditions. The selected time windows can be seen in Figure 3. The analysis of the two ERPs is performed within-participant and replicated across participants. The statistical tests are performed over the trials that each participant goes through, divided into the three image categories, giving 160 trials per category and participant.

The first hypothesis is that the N170 amplitude will be stronger (more negative, since the component has a negative pole) in the animated condition, but not in the semi-realistic condition when compared to the human condition. From a visual inspection of the ERP curves in Figure 4, it is immediately apparent that there is a difference between the categories in the N170. This is also confirmed through an independent sample t-test, as characters belonging to the *unrealistic* category elicit a significantly ($p < 0.05$) stronger N170 than characters in the *realistic* category in 4 out of 5 participants (all but participant 'd').

The second hypothesis is that *semi-realistic* characters will be the only ones eliciting an N400 response representing as their later cognitive processing contradicts the earlier human face processing. This difference is apparent in the ERP curves in Figure 4 and an independent sample t-test shows a statistically significant difference in the intensity of the N400 component ($p < 0.05$) for the *semi-realistic* compared to the *realistic* characters in 4 out of 5 participants (all but participant 'a'). We do not find significant evidence for an N400 response instead when analysing the response to *unrealistic* characters.

⁵The dataset, containing the collected data and the stimuli, is available at <https://doi.org/10.5281/zenodo.7948158>

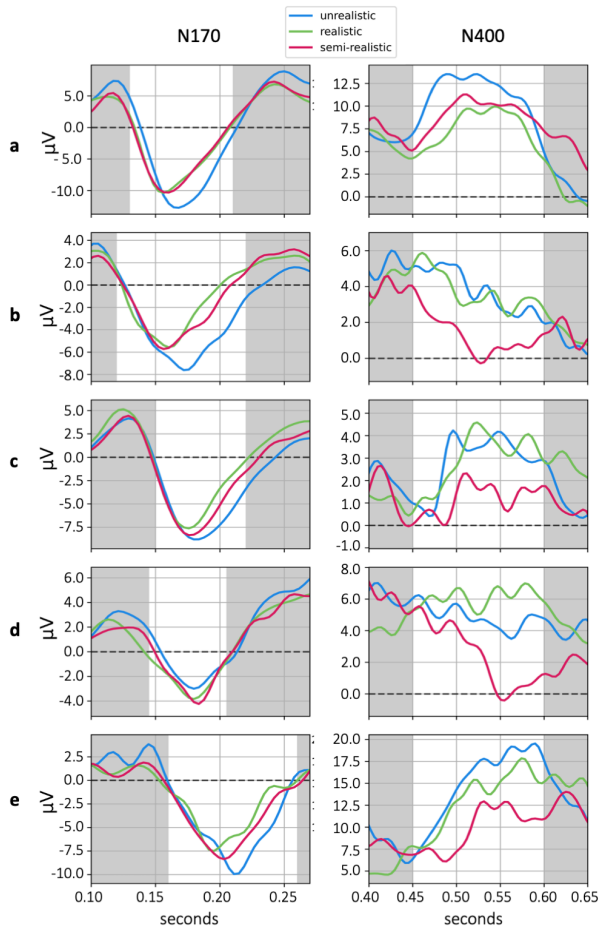


Fig. 4: ERP waves in the N170 and N400 time windows for the three categories. Each row corresponds to one participant.

unrealistic chars not of interest

V. DISCUSSION

The results of the ERP analysis show that, in 4 out of 5 participants, the *unrealistic* characters elicit a stronger N170 face effect when compared to the *realistic* ones. In contrast, only one participant demonstrates the same increased N170 effect in the *semi-realistic* condition. This suggests that at this stage of processing facial stimuli, the *unrealistic* faces are processed differently than the *realistic* and *semi-realistic* faces. The N170 is considered to mark the brink of conscious processing of facial stimuli and not the immediate recognition of a face [22]. At this stage of processing, the participant is discerning the high-level characteristics of the face, such as sex, age, and expression, based on individual features. It seems, therefore, that the unusual composition of the unrealistic characters engages the face-specific brain processing more, as it might be harder to discern high-level characteristics of the character. However, this also suggests that the UV phenomenon does not stem solely from this step in facial processing, since the unrealistic and semi-realistic characters chosen from the survey were specifically chosen to elicit the lowest and highest uncanniness response, respectively.

By looking at the later event-related potential response, we

can see that 4 out of 5 participants have a significantly stronger N400 response to *semi-realistic* characters when compared to the *realistic* one and no participant had a corresponding response to the *unrealistic* characters. This suggests that some form of semantic error correction is happening when looking at these almost-but-not-quite-human faces. As suggested by the N170 analysis, the early processing of semi-realistic faces resembles the processing of human faces due to the similar proportions. But as the participant realizes the face is not actually a human face, an error-correcting response is necessary, as the character is indeed not an actual human. This interpretation is in line with the prediction of the dehumanization hypothesis [34] as an expectation of animacy and therefore humanity is generated in the early stage of visual processing, which triggers an error correction afterwards, as one realizes the character lacks actual inner life. We suggest the N400 response marks the beginning of the error-correcting mechanism of this dehumanizing process.

In both ERP analyses, one of the five participants did not follow the pattern of the other four. In group-level analyses with a higher number of participants, these incongruities would go unnoticed if the effect was strong enough. However, in a small N study like ours, where each participant essentially functions as a replication, deviant results warrant a discussion. In the N170 analysis, the participant 'd' is the outlier. Looking closer at the participant's N170 (Figure 3 and 4), one can see the amplitude of the N170 has a lower amplitude and a shorter time-window compared to other participants. This observation might be the consequence of the single-electrode nature of this measurement or might be the manifestation of an inter-participant difference. We believe further studies would be necessary to verify the origin of this ERP outlier.

In the N400 analysis another participant, participant 'a', does not show a significantly negative peak in the later time window. Based on the rating of animacy collected from each participant, we can say that this person rarely gave a medium score on animacy regardless of realism, particularly in the *semi-realistic* condition, when compared to other participants. The lack of an N400 effect could stem from this participant simply not perceiving the *semi-realistic* characters as resembling humans. A solid link between behavioural and neurological results cannot be established from this single participant alone, but it provides a possible reconciliation of the result's deviancy from the rest of the group. To draw a clearer picture of the individual differences in terms of UV response, future studies in this direction should capture individual behavioural responses contextually to the EEG measurements. A larger-scale study including this kind of data has the potential to lead to a better understanding of the link between subjective reports and EEG signals.

On a final note, this experiment follows a well-established protocol to capture event-related potential neural responses, which requires long sessions with a large number of stimuli repetitions to achieve the best signal-to-noise ratio. However, the elevated robustness and validity of the measurements given by this design come at the expense of a potentially

lower ecological validity, as there is a clear gap between the experimental condition and the real context in which a player would interact with digital characters. Bridging this gap, with more realistic experimental protocols, could unlock a more widespread usage of neurophysiological data in player experience modelling, leading potentially to more accurate and empirically-grounded models.

VI. CONCLUSIONS

In this study, we investigate the cognitive origins of the Uncanny valley phenomenon by collecting behavioural and EEG responses to a set of images in a survey and a laboratory experiment. Through the survey, we select a set of images that elicit a behavioural response compatible with the Uncanny Valley, these images are used in the EEG experiment to capture the neural response of the participants at different locations and time frames. The data and the analysis give a strong indication that the phenomenon is a consequence of a mismatch in the early and later cognitive processing, in which an uncanny face is first detected as human and later “corrected”. While a number of open questions still remain to be answered, especially in terms of individual differences between the participants and how to study them, we believe the results contribute to a better understanding of the Uncanny Valley phenomenon and showcase the applicability of neurophysiological studies to player experience analysis.

REFERENCES

- [1] Milena T. Bagdasarian, Anna Hilsman, Peter Eisert, Gabriel Curio, Klaus-Robert Müller, Thomas Wiegand, and Sebastian Bosse. EEG-Based Assessment of Perceived Realness in Stylized Face Images. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–4, May 2020. ISSN: 2472-7814.
- [2] Nima Bigdely-Shamlo, Tim Mullen, Christian Kothe, Kyung-Min Su, and Kay A. Robbins. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, 9, June 2015.
- [3] Timothy A. Brown. *Confirmatory Factor Analysis for Applied Research, Second Edition*. Guilford Publications, January 2015. Google-Books-ID: tL2BQAAQBAJ.
- [4] Marcus Cheetham, Pascal Suter, and Lutz Jäncke. The Human Likeness Dimension of the “Uncanny Valley Hypothesis”: Behavioral and Functional MRI Findings. *Frontiers in Human Neuroscience*, 5, 2011.
- [5] Marcus Cheetham, Lingdan Wu, Paul Pauli, and Lutz Jäncke. Arousal, valence, and the uncanny valley: psychophysiological and self-report findings. *Frontiers in Psychology*, 6, 2015.
- [6] Lisa DeBruine and Benedict Jones. Face Research Lab London Set, 2017. Artwork Size: 88208545 Bytes Pages: 88208545 Bytes.
- [7] Arnaud Delorme, Jason Palmer, Julie Onton, Robert Oostenveld, and Scott Makeig. Independent EEG Sources Are Dipolar. *PLoS ONE*, 7(2):e30135, February 2012.
- [8] Alexander Diel, Sarah Weigelt, and Karl F. Macdorman. A Meta-analysis of the Uncanny Valley’s Independent and Dependent Variables. *ACM Transactions on Human-Robot Interaction*, 11(1):1:1–1:33, October 2021.
- [9] Alexandre Gramfort. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7, 2013.
- [10] Chin-Chang Ho and Karl F. MacDorman. Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, 26(6):1508–1518, November 2010.
- [11] Mainak Jas, Denis A. Engemann, Yousra Bekhti, Federico Raimondo, and Alexandre Gramfort. Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, 159:417–429, October 2017.
- [12] Herbert H Jasper. The ten-twenty electrode system of the International Federation. *Electroencephalogr. Clin. Neurophysiol.*, 10:370–375, 1958.
- [13] Emily S. Kappenman and Steven J. Luck. ERP Components: The Ups and Downs of Brainwave Recordings. December 2011.
- [14] Marta Kutas and Kara D. Federmeier. Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual review of psychology*, 62:621–647, 2011.
- [15] Jari Kätsyri, Beatrice De Gelder, and Tapio Takala. Virtual Faces Evoke Only a Weak Uncanny Valley Effect: An Empirical Investigation With Controlled Virtual Face Images. *Perception*, 48(10):968–991, October 2019.
- [16] Tanja Lischetzke, David Izydorczyk, Christina Hüller, and Markus Appel. The topography of the uncanny valley and individuals’ need for structure: A nonlinear mixed effects analysis. *Journal of Research in Personality*, 68:96–113, June 2017.
- [17] Karl F. MacDorman and Debaleena Chattopadhyay. Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition*, 146:190–205, January 2016.
- [18] Karl F. MacDorman and Debaleena Chattopadhyay. Categorization-based stranger avoidance does not explain the uncanny valley effect. *Cognition*, 161:132–135, April 2017.
- [19] Masahiro Mori, Karl MacDorman, and Norri Kageki. The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine*, 19(2):98–100, June 2012.
- [20] Maryam Mustafa, Stefan Guthe, Jan-Philipp Tauscher, Michael Goesele, and Marcus Magnor. How Human Am I?: EEG-based Evaluation of Virtual Characters. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 5098–5108, Denver Colorado USA, May 2017. ACM.
- [21] Astrid M. Rosenthal-von Der Pütten and Nicole C. Krämer. How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior*, 36:422–439, July 2014.
- [22] Bruno Rossion. Understanding face perception by means of human electrophysiology. *Trends in Cognitive Sciences*, 18(6):310–318, June 2014. Publisher: Elsevier.
- [23] Bruno Rossion and Corentin Jacques. The N170: Understanding the time course of face perception in the human brain. In *The Oxford handbook of event-related potential components*, Oxford library of psychology, pages 115–141. Oxford University Press, New York, NY, US, 2012.
- [24] Kyoshiro Sasaki, Keiko Ihaya, and Yuki Yamada. Avoidance of Novelty Contributes to the Uncanny Valley. *Frontiers in Psychology*, 8:1792, October 2017.
- [25] Sebastian Schindler, Eduard Zell, Mario Botsch, and Johanna Kissler. Differential effects of face-realism and emotion on event-related brain potentials and their implications for the uncanny valley theory. *Scientific Reports*, 7(1):45003, March 2017.
- [26] Mincheol Shin, Se Jung Kim, and Frank Biocca. The uncanny valley: No need for any further judgments when an avatar looks eerie. *Computers in Human Behavior*, 94:100–109, May 2019.
- [27] Philip L. Smith and Daniel R. Little. Small is beautiful: In defense of the small-N design. *Psychonomic Bulletin & Review*, 25(6):2083–2101, December 2018.
- [28] Avinash Tandle, Nandini Jog, Pancham D’cunha, and Monil Chheta. Classification of Artefacts in EEG Signal Recordings and EOG Artefact Removal using EOG Subtraction. In *Communications on Applied Electronics*, volume 4, pages 12–19, January 2016. ISSN: 23944714 Issue: 1 Journal Abbreviation: CAE.
- [29] Angela Tinwell. *The Uncanny Valley in Games and Animation*. CRC Press, December 2014. Google-Books-ID: 8mZYBQAAQBAJ.
- [30] Angela Tinwell, Mark Grimshaw, Debbie Abdel Nabi, and Andrew Williams. Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Computers in Human Behavior*, 27(2):741–749, March 2011.
- [31] Burcu A. Urgan, Marta Kutas, and Ayse P. Saygin. Uncanny valley as a window into predictive processing in the social brain. *Neuropsychologia*, 114:181–185, June 2018.
- [32] Julija Vaitonytė, Maryam Alimardani, and Max M. Louwerse. Scoping review of the neural evidence on the uncanny valley. *Computers in Human Behavior Reports*, 9:100263, March 2023.
- [33] Shensheng Wang, Yuk F. Cheong, Daniel D. Dilks, and Philippe Rochat. The Uncanny Valley Phenomenon and the Temporal Dynamics of Face Animacy Perception. *Perception*, 49(10):1069–1089, October 2020.
- [34] Shensheng Wang, Scott O. Lilienfeld, and Philippe Rochat. The Uncanny Valley: Existence and Explanations. *Review of General Psychology*, 19(4):393–407, December 2015. Publisher: SAGE Publications Inc.