The face-eye mismatch hypothesis:

The role of a mismatch in face and eye region in the emergence of the UV effect

Jonas Schlüter

Master Thesis Human Factors & Engineering Psychology

April 30, 2024

First Supervisor: Dr. Martin Schmettow

Second Supervisor: Dr. Cesco Willemse

Faculty of behavioral, Management and Social Sciences

University of Twente

Abstract

Since the discovery of the Uncanny Valley effect, the feeling of creepiness in response to near human-like stimuli, researchers have been looking for its cause. The present study investigates the face-eye mismatch hypothesis, which states that a mismatch between the face and eye region results in the eerie feeling. Thus, causing the UV effect. To test this hypothesis a new stimulus set was created with the help of AI generation. In the current study, 26 participants were presented with congruent and incongruent human and primate stimuli. Congruent faces were not manipulated. Incongruent faces have an altered eye region that consists of a darkened sclera for human stimuli and a set of human eyes for primate stimuli. Congruent human faces were rated as more likeable compared to incongruent human faces. Primate congruent faces were rated more likeable than their incongruent counterparts, although the difference was much smaller than in human stimuli. Thus, our results support the face-eye mismatch hypothesis. The sub-hypothesis stated that oscillation count is increased for incongruent stimuli. The oscillation count was slightly lower for primate stimuli than for human stimuli. However, the oscillation count did not differ across congruency conditions for both species. Therefore, the findings do not support the theory that oscillation counts can be used as proxy measurement for eerieness.

Keywords: uncanny valley, eye tracking, oscillation

Table of Contents

Introduction

Public significance

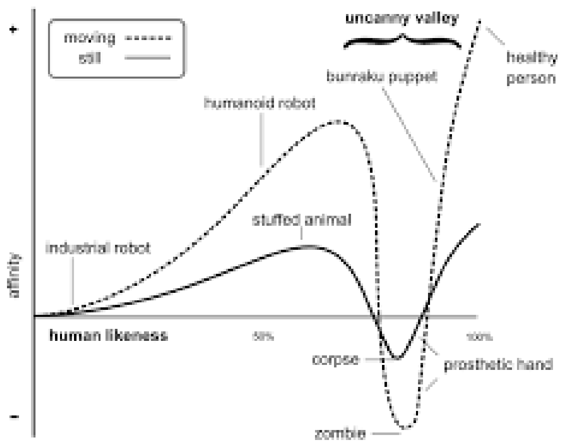
Computer-generated imagery (CGI) characters have come a long way since their humble beginnings in movies like Tron or Toy story. With the remarkable success of Marvel superhero movies, CGI Characters such as Groot have solidified their importance for the modern entertainment industry. The use of CGI avatars has become more widespread for content creators to represent themselves in their broadcasts to thousands of viewers. In addition to the entertainment industry, robotics experts are working on ever more capable and human-like robots to aid in the care of the elderly, the household or in rescue missions. However, even before the first film using CGI was published Dr Masahiro Mori characterized a phenomenon in 1970, which shapes our acceptance of machines and artificial characters to this day, the Uncanny Valley (UV) effect.

The Uncanny Valley Effect

Mori (1970) stated that attaching human-like features to machines and robots will result in a heightened sense of affinity. However, he noticed that stimuli which reach near-perfect resemblance to humans, can induce an uncanny and eerie feeling which results in a negative emotional response. This negative emotional response is again reversed if the human-likeness of the stimuli is increased further (Figure 1). Since then, Mori’s observation has been replicated numerous times by independent researchers (Burleigh, Schoenherr & Lacroix, 2013; MacDorman & Chattopadhyay, 2016; Keeris & Schmettow, 2016). As the effect's existence has been established, many theories have been put forth on how it emerges.

Figure 1

The Uncanny Valley Effect Curve by Mori (1970)



Note. This graph depicts the affinity an observer associates with certain stimuli dependent on their degree of human likeness.

An early theory, called the mortality salience hypothesis, states that replications of a human body remind the observer of their own mortality (MacDorman & Ishiguro, 2006). This reminder results in fear for their own lives, which elicits an eerie feeling and negative emotional response. It is even argued that humans feel threatened by androids and fear being replaced by them (Ho et al., 2008). A different approach, called the perceptual mismatch hypothesis, suspects inconsistencies in the object features' realism level to be the cause for the severe drop in likeability observed in the UV effect (Pollick, 2010). Whereas the categorization ambiguity hypothesis assumes there is ambiguity when categorizing a stimulus as human or non-human. Stimuli that elicit the UV effect would cause higher perceptual discrimination difficulty (Cheetham et al., 2014). According to MacDorman, Green, Ho, & Koch (2009) the effect is an evolutionary mechanism that helps with disease/threat avoidance and mate selection. MacDorman (2009) et al. differentiate between fast- and slow-systems explanations for this effect. Fast-systems explanations stemming from automatic, stimulus-driven processing early in perception. Whereas slow system explanations assume conscious deliberation processes occurring later in cognitive processing to cause the effect. While Keeris (2016) was unable to show the effect with a presentation time of 50 ms, Haeske (2016) managed to replicate the uncanny valley curve with a stimulus presentation time of 100 ms. In addition to that, Slijkhuis (2017) showed that 50 ms are sufficient to reliably judge the eeriness of a face for most participants. This lends strong support to the fast-system theories.

Such theories include the threat/disease avoidance theory and evolutionary aesthetics. Moosa & Ud-Dean (2010) hypothesize that the uncanny valley effect stems from an evolutionary danger avoidance mechanism rather than disease avoidance. A study by Koopman and Schmettow (2019) tested the universality of the uncanny valley effect. The results indicate the uncanny valley effect to be a universal experience across all participants. Which suggests that the uncanny valley effect is an innate cognitive mechanism rooted in evolution (Stikker, 2023). Furthermore, Geue and Schmettow (2021) managed to show the uncanny valley effect applies not only to robots and CGI characters but also to biological faces, varying in their evolutionary closeness to the human. Here, more human likeness or evolutionary closeness resulted in increased perceiver's affinity. This lends further support to an evolutionary origin of the effect.

Importance of the eye region and eye tracking

In search for the evolutionary origin of the effect, Limmer (2023) tested the scleral-facial mismatch theory as a cause for uncanny feelings, but his findings did not fully support the theory. His findings suggest that participants overall would give more attention to the eye region compared to nose and mouth. In addition to that, visual attention was higher in the eye region of mismatched faces compared to congruent ones. This suggests that certain areas of the face, especially the eyes, have a bigger influence than others in determining human likeness. Prior studies indicate that, when humans interact, the eye region is most often attended out of all facial features (Itier and Villate, 2007;). Regardless of the nature of the interaction, the eyes remain the preferred source of information (Itier and Batty, 2009). This makes sense as all basic facial expressions of feelings are recognizable by a part of the eye region (Schyns et al., 2002). Even more complex feelings, e.g. guilt and envy, have been shown to be recognizable by looking at the eye region alone (Baron-Cohen et al.; 1997; Baron-Cohen et al., 2001). Furthermore, research by Bagepally (2015) shows that participants in most eye-tracking studies report a quick fixation of the eye region when a face is presented to them. This also indicates a high priority of the eye region for face processing tasks.

Another study, by Stikker (2023), sought to examine the specific category confusion theory using human, animal and incongruent (human with animal eye or vice-versa) stimuli as well as sunglasses to cover the eyes. The results supported the hypothesis that animal faces with human eyes were perceived as least likeable. However, a clear support for the specific category confusion theory could not be found. The uncanny animal (with human eyes) group did indeed get higher likability scores when pictured with sunglasses, however, this was the same for regular animals. This led the researcher to conclude that the eye region is important in the process but not the sole cause of the uncanny valley effect.

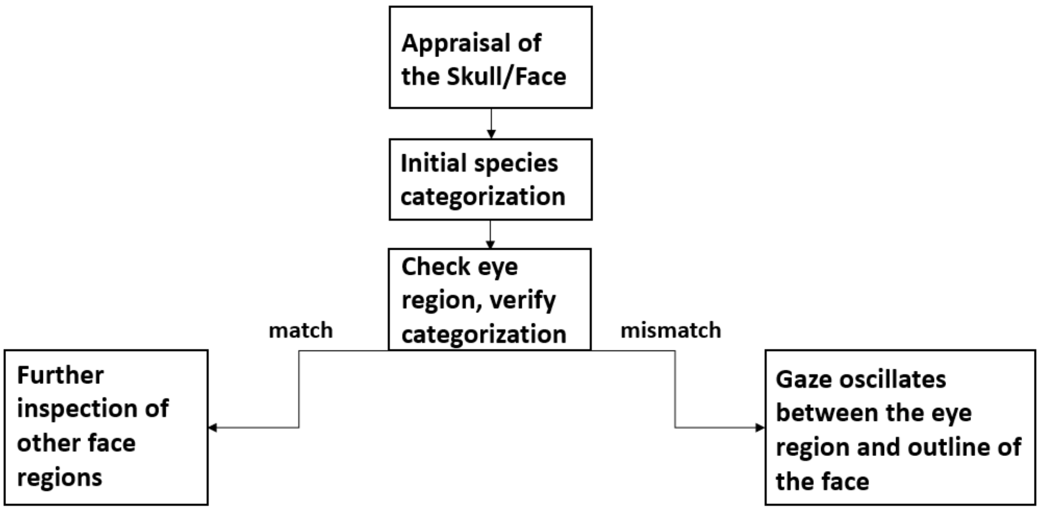
While the studies of Limmer and Stikker failed to produce strong support for the proposed theories the findings are in line with previous eye-tracking research by Grebot et al. (2021) regarding the difference in importance of different regions of interest (eyes, nose, and mouth). Grebot et al. (2021) found that difficulties in perceptual discrimination in faces that lie on the border of being an Avatar or human-like, would change the allocation of attention. According to the mind-eye hypothesis, what the mind processes is reflected by eye movements (Beesley et. al, 2019). Therefore, eye-tracking methods could provide more insight into the cognitive processing of stimuli that elicit uncanny feelings (Cheetham et al., 2013).

Current study

This section aims to elaborate how prior research shapes our hypothesis about explanations for the UV effect. As mentioned in the prior sections, social cognition research indicates that the UV effect is a fast-system mechanism with evolutionary origin. Based on earlier research, we expect the eyes to be a crucial factor in evaluating the human-likeness of a stimulus. As the human eye differs from other primates in respect to coloration of the sclera, it is a suitable criterion to discern an individual's species belongingness. Additionally, the eye region can deliver a lot of information about a person's current emotional state. For example, a fake smile can be distinguished from a genuine one, by examining wrinkles around the eye characteristic for a genuine laugh (Baron-Cohen et al., 2001). The ability to gauge an individual's species belongingness or emotional state carries with it many evolutionary benefits in the context of threat avoidance and evolutionary aesthetics (Mori et al., 2012). In the context of threat avoidance, an individual who does not belong to your species or expresses hostility poses a potential threat. In the context of evolutionary aesthetics, it would aid in mate-selection and identification of potential within-species rivals. However, it is highly unlikely that the eye region is the sole factor when making these assumptions, as throughout history we would have encountered other humans and hominids. Multiple types of humans lived concurrently, namely the early Homo sapiens, the Homo neanderthalensis, the Denisova hominis and the Homo naledi. This would have made species identification by eye region alone a difficult endeavor to say the least. This, in combination with earlier research, leads us to assume that the eye region acts as a control mechanism for an assumption. This assumption is based on the appearance and shape of the whole face. Prior studies regarding face processing have found that faces are holistically processed by humans. This means, an overall “gestalt” of a face is processed before the individual regions that it is comprised of (Maurer et al., 2002). We argue that hierarchical scanning takes place where a first look at the skull leads to an assumption whether any given individual belongs to one's species. Only then, is the eye region examined to validate the prior assumption. A mismatch between the eye region and the prior assumption based on the face would result in an eerie feeling. This idea is in line with other theories that see the UV effect as a corrective mechanism that activates whenever the brain receives conflicting signals during species categorization. Contrary to other studies, we hypothesize a hierarchical gaze order. We assume this hierarchy to start with an initial evaluation of the face or skull shape which leads to an assumption regarding the species categorization. Subsequently, this assumption is verified by glancing at the eye region. A mismatch between the eyes and the human-like appearance of the skull would prompt a quick re-evaluation of both stimuli. This would result in quick succession of gazes, or oscillations, flickering between the skull and the eye region (see Figure 2). We call this theory the Face-Eye mismatch theory. It would be helpful for future research to establish a scan pattern for the UV effect as this would provide an addition to the commonly used likeability scale (Mathur & Reichling, 2016) to measure if certain stimuli elicit the UV effect.

Figure 2

Schematic of the expected gaze patterns.



Note. Schematic shows which gaze pattern is expected during the first 2500 ms when encountering a depiction of an unknown face. A mismatch in initial categorization and the eye region verification elicits a feeling of uncanniness.

In the current study we plan to test this theory on a new set of stimuli by using Eye-tracking equipment and software to investigate the number of oscillations between areas of biological faces. This is done in addition to commonly used self-reported likeability rating scales. Due to the problems in prior research with animal stimuli (Stikker, 20023) and the newly emerged possibilities of image generation software we decided to create a new stimulus set. A detailed description of the stimulus set creation process is given in the methods sections below. We will use congruent faces (humans and primates) and incongruent faces (Humans with primate eyes and vice versa) as stimuli.

We expect congruent human and primate faces to not cause feelings of uncanniness. They would therefore be perceived as likeable by the participants and rated highly. However, incongruent faces, meaning human faces with primate eyes and primates with human eyes, are expected to elicit feelings of uncanniness. Thus, resulting in less perceived likeability and lower ratings. Primate faces with human eyes are expected to be rated as less likeable than human faces with primate eyes based on findings by Stikker (2023).

Research Question and Hypothesis

Does the Face-eye mismatch cause the uncanny valley effect?

H1 Congruent stimuli, faces in which skull shape and eye region match, do not cause feelings of uncanniness.

H2 Incongruent stimuli, faces in which skull shape and eye region do not match, do not cause an eeriness response.

H3 Incongruent primate faces, primate faces with a human eye region, cause stronger feelings of uncanniness than incongruent human faces

In addition to that, we want to test whether incongruent stimuli cause participants gazes to oscillate more frequently compared to congruent stimuli. Looking at faces that elicit the UV effect, we expect participants to evaluate the general shape or “gestalt” first, then move on to the eye region and oscillate between both afterwards to confirm their initial evaluation. Such a unique oscillation pattern could be used as a behavioral measurement for the UV effect and would be a great addition to self-reported likeability scales. It could essentially be used as a proxy measure. To inform the current study and gain insight into prior UV effect research that uses eye tracking metrics, a systematic literature review was conducted.

Uncanny Valley Effect Research using Eye-Tracking Technology:

A systematic Literature Review

Literature Review: Methods

Definition of Terms

The uncanny valley (UV) effect describes a dip in the observer's affinity for a given object when approaching near perfect human likeness – a valence that otherwise increases with the object's human likeness. The object in question can be a biological face but it is most often artificial, e.g., computer-generated images, robots, drawn faces or even prosthetics.

Oscillations refer to rapid eye movements between two objects of interest. It does not refer to the medical term which describes an involuntarily movement of the eyes away from the object of interest.

Research Questions

This review aims to identify the methods and possible shortcomings of eye tracking studies concerning the UV effect. Specifically, we are interested in the prior use of oscillations as an eye tracking metric.

R1 What type of eye tracking variables are used to investigate the UV effect?

R2 What is the theoretical reasoning behind using a specific variable?

R3 Have oscillating eye movements been used before as a proxy measure?

Eligibility Criteria

Only articles were included that were:

I. Listed in the Scopus database
II. Published in a peer reviewed journal and written in English
III. Based on an experimental study using eye tracking technology to research aspects of the UV effect

Articles that do not meet the criteria are excluded.

Search Strategy

The string used to search the Scopus database was:

TITLE-ABS-KEY (“uncanny AND valley AND effect, AND eye AND tracking”)

Literature Review: Results

A total of 13 documents were found in the initial search. Two of these documents contain results of a study conducted with non-human primate stimuli non-human primates. These studies are still deemed relevant as similar stimuli are used in the current study. One study used eye tracking to improve the creation of personalized avatars. However, the study did not investigate the UV effect in an experimental design and was therefore excluded. The same applies for a study using eye tracking to investigate usability testing in Virtual Reality. A study investigating the salience of anomalies in animated human characters was excluded as it did not use eye tracking. Furthermore, the study is focused on the difference between body and face anomalies while the current study investigates the impact of a mismatch in facial features. Additionally, one of the documents was not available as full text and the abstract did not mention any eye tracking methods used in an experimental design. Therefore, it was excluded. See Table 1 for an overview of all included publications.

Table 1

Overview of included journals

Title of paper	**Authors**	**Year of Publication**	**Topic of paper**	**Eye tracking metrics**
The other-race effect in the uncanny valley	A. Saneyoshi, M. Okubo, H. Suzuki, T. Oyama, & B. Laeng	2022	The other-race bias in UCVE – Participants rated same race UCVE triggering stimuli more unpleasant	Diameter of the pupil
Accepting Humanlike Avatars in Social and Professional Roles	M. Sharma, & K. Vemuri	2022	3 experiments regarding the acceptance of avatars. Ambiguous response in more complex situations	face, eye region, jaw and mouth, and torso AOI. pupil size variation, fixation count
Infant discrimination of humanoid robots	G. Matsuda, H. Ishiguro, & K. Hiraki	2015	An android is used to examine infant discrimation ability between human and robots	AOIs (Face, goal, body. Time spent on AOI

Category processing and the human likeness dimension of the uncanny valley hypothesis: eyetracking data	M. Cheetham, I. Pavlovic, N. Jordan, P. Suter, & L. Jancke	2013	Forced choice categorization task with gradually morphed human/avatar faces	Number of fixations, dwell time
Uncanny Valley Hypothesis and Hierarchy of Facial Features in the Human Likeness Continua: An EyeTracking Approach	I. B. da Fonseca Grebot, P. H. P. Cintra, E. F. F. de Lima, M. V. de Castro, & R. de Morae	2022	Investigating the effect of perceptual ambiguity on the hierarchical processing of facial features	Gaze & dwell time on eye, nose, and mouth AOI
Macaque Gaze Responses to the Primatar: A Virtual Macaque Head for Social Cognition Research	Wilson, V.A.D., Kade, C., Moeller, S., ...Kagan, I., Fischer, J.	2020	Development of a virtual avatar with varying degrees of realism. No evidence for UV effect was found	Total time spent
A naturalistic dynamic monkey head avatar elicits species-	Siebert, R., Taubert, N., Spadacenta, S.,	2020	Investigating the reaction of rhesus macaque to rhesus monkey avatar with	Total time spent

typical reactions and overcomes the uncanny valley	...Giese, M.A., Thier, P.		varying degree of naturalism

R1 What type of eye tracking variables are used to investigate the UV effect?

The two studies conducted with primate stimuli and participants only used the total time spent looking at a certain stimulus as a metric. Four of the seven papers used an AOI based metric. Two of those papers used gaze and dwell time on specific AOIs. The total number of fixations was also used by two papers out of the seven. The pupil diameter variation, meaning the change in pupil size, was used by two studies.

R2 What is the theoretical reasoning behind using a specific variable?

In the two studies using non-human primates, the total time spent looking at a stimulus is used together with behavioral signs of arousal, for example facial expressions or general motion, to make inferences about the affective response of the primate. A difference in looking times would count as evidence for the UV effect. One study found a difference that indicates the existence of the UV effect in monkeys (Siebert et al., 2020). Matsuda et al. (2015) also used something they called “looking times”. Looking times were calculated by dividing the mean gaze count by the total gaze count for each of the three age groups/ stimuli. They noted that at a recording requency of 300 hz, one gaze is comprised of 3.3ms. The study by Saneyoshi et al. (2022) used variation in the pupil diameter as a measure for the affective response of participants to a stimulus. They argued that this would be an objective measure in addition to the common rating scales for likeability. In their first two experiments, Sharma and Vemuri (2022) use this metric with the same reasoning. The third experiment of this study used the number of fixations on the eye, nose and mouth AOI. Similar AOIs were used by Cheetham et al. (2023) to investigate the hierarchical order of these AOIs in face processing. Here, dwell time was used in addition to the fixation count. The Researchers consciously chose to limit the amount of eye tracking metrics to prevent data dredging. The paper by da Fonseca Grebot et al. (2022) uses a similar approach and builds on the work by Cheetham et al. (2013). However, they use gaze dwell time per AOI instead of fixation-based dwell time. This is done because the parameters of fixations are not pre-defined and could lead to varying measurement results.

R3 Have oscillating eye movements been used before as a proxy measure?

None of the papers found in the literature research mention the use of oscillations between different regions of the face as an indicator for categorization difficulties. However, the need for more objective measures than self-reported rating scales was mentioned in prior studies (Saneyoshi et al., 2022). However, pupil dilation was used and mentioned as a possible proxy measure for eeriness (Saneyoshi et al., 2022; Sharma & Vemuri, 2022).

Additional Research Question based on literature review

Through the literature review's findings, a need for more objective measurements of the UV effect became apparent. Saneyoshi et al. (2022) and Sharma and Vemuri (2022) try to incorporate more objective measures by using pupillary dilation as an indicator for surprises and the nature of an emotional response. Although not mentioned in prior research, the oscillations of the gaze back and forth from the eye region could also prove to be an additional objective measurement for the UV effect. We argue that stimuli triggering the UV effect would result in a higher number of gaze oscillations in the same way that an increase in pupillary diameter can be observed. Based on these findings a sub hypothesis for the current study can be formulated.

H4 Incongruent stimuli, faces in which skull shape and eye region do not match, increases the number of gaze oscillations to the eye region.

Methods

Participants

The current study used a convenience sample of 26 students from the University of Twente. Of the 26 students, 14 were female, and 12 were male. Students used Sona Systems, a web-based test subject pool service, to sign up for participation. Students receive Sona credits for participation in the study. Participants with impaired vision had to be excluded from the study, since wearing glasses or contact lenses can interfere with data collection using eye tracking devices.

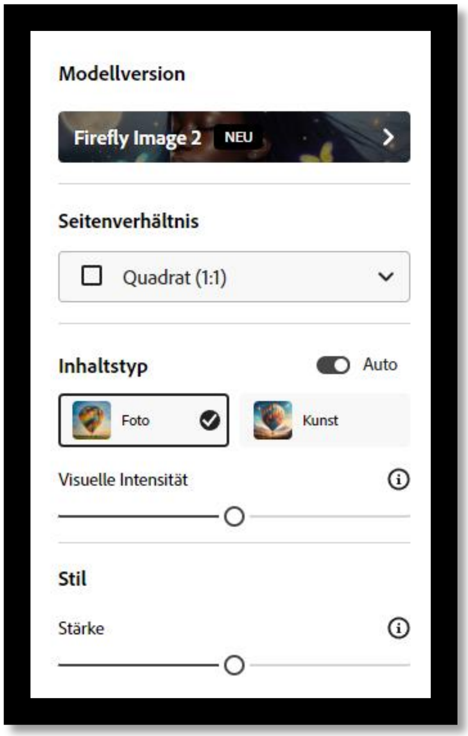
Materials

Stimuli

The stimuli set from the University of Twente, used in prior research by other students contains a lot of pictures that are not optimally suited to be used in an experiment. Some have low resolution, others depict faces with extreme facial expressions of anger or sadness. This could influence the likeability of the stimulus. Therefore, we aimed to create a new stimuli set with higher resolution pictures and neutral facial expressions. The new stimuli set features several real-life pictures of humans without any extreme facial expressions. Optimal pictures of primates, that meet those requirements, are harder to obtain. In recent years, highly naturalistic generated primate stimuli have been shown to overcome the UV effect (Siebert et al., 2020). Therefore, AI generation is used to create primate stimuli.

Figure 3

Settings for Adobe Firefly "text to picture" generation of primates



Note. Visual intensity, Style and strength remained the same. The type of content is set to Foto instead of Art to ensure a realistic picture.

The primate stimuli were generated using text prompts in the Adobe Firefly software "text to picture" function. The settings were left to standard except for the type of content, which was set to Photo instead of Art (see Fig. 3). This was done to ensure that we generate photorealistic pictures that do not deviate from another in style, orientation, angle or lighting. The first text prompt used was "face and head of a primate". The word primate was altered to ape, orangutan, chimpanzee, Lemur, and Gorilla to obtain pictures of various kinds of primates (see Fig. 4). The output of the software always consists of 4 pictures. Not all the results are fully convincing as real-life primates due to several reasons. For example, unnatural fur or eye coloration and odd or weird facial expressions. Therefore, right after picture generation an initial selection is made by the researcher to select the most realistic and suitable stimuli. One or two pictures are always excluded due to the orientation of the head and eyes. In total 20 primate pictures made it through the selection process and were included in the study.

Figure 4

Primate stimuli from the current study

Note. The prompt "Gorilla" was used for the left picture, "Chimpanzee" on the right.

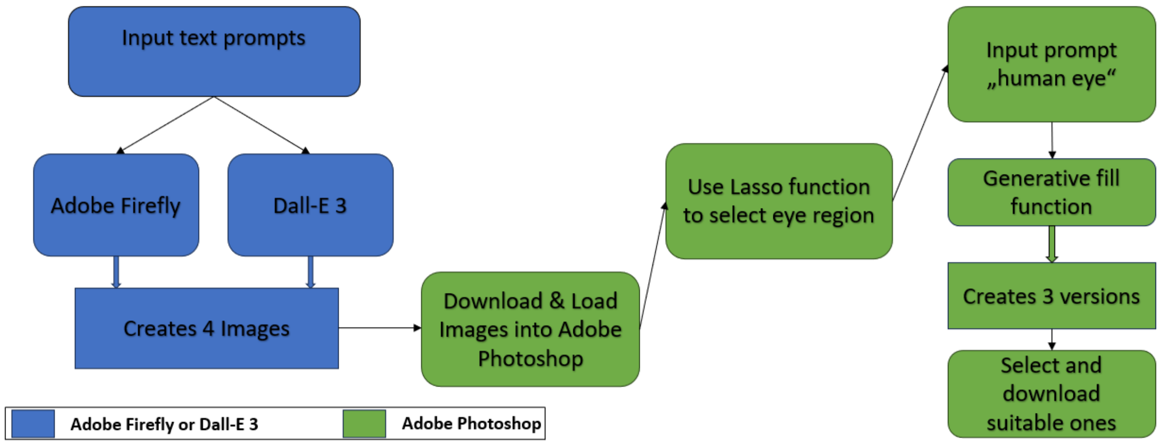
Stimuli containing Human faces were taken from the Chicago Face Database (CFD). It offers multiple dozen pictures of people with varying ethnic backgrounds around the Chicago area in the USA. It was compiled as a free stimulus set of faces by Ma, D. S., Correl, J., & Wittenbrink, B. in 2015. A selection of 22 pictures was made by the researchers to match the amount of primate stimuli and keep the set manageable. Criteria for the selection process were a neutral facial expression, multiple ethnicities, and an equal amount of male and female individuals.

Manipulation

To create incongruent primate stimuli the "generative fill" function in Adobe Photoshop is used (see Fig. 5). The primate pictures generated in Adobe Firefly or Dall-E 3 are loaded into photoshop after selection. The eye region of the primates is selected using the "Lasso" function. Using the Input prompt "human eye" in the "generative fill" function results in 3 versions of an altered eye region that resembles that of a human.

Figure 5

Workflow for primate stimuli creation and manipulation.



Note. Steps for stimuli creation that were done in Adobe Firefly are marked blue. The green steps were done in Adobe Photoshop.

The technology is new and has some flaws. To always generate a human pair of eyes with white sclera some pictures must be prepared for the generative fill function. This is done by drawing a white line into the eye of the original before proceeding with the aforementioned steps. Only the most natural looking one of the three versions is selected and saved as our incongruent primate stimuli. See Figure 6 for some examples. This is done for all 20 AI generated primate stimuli.

Figure 6

Manipulated primate stimuli from the current study

Note. Stimulus generation in Adobe firefly results in three pictures. Only the stimuli with the most human like eye region are chosen as final stimuli.

In order to manipulate the human stimuli in our set a copy is made of each stimulus and changed manually as the generative fill function yields poor results when attempting to generate the eye region of a primate on a human face. To let the human eye region appear more primate like their scleral coloring will be darkened. See Figure 7 for some examples. This method was used in prior research by Limmer (2023) and yielded satisfactory results. A visual design expert was consulted to inform the manipulation process.

Figure 7

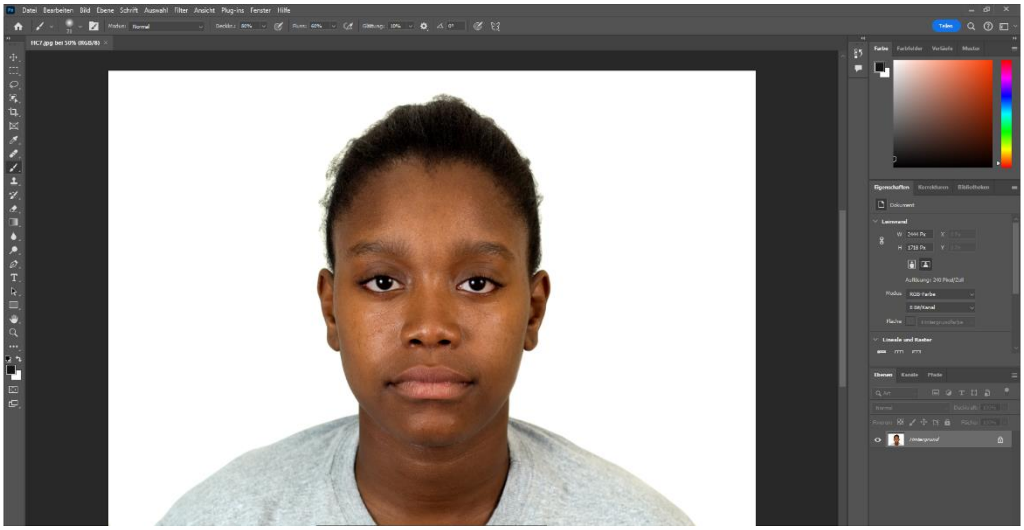
Examples of manipulated human stimuli from the present study.

Note. This figure depicts a male and female human face. The sclera has been darkened to incongruent stimuli.

The air brush tool in Adobe photoshop was used to gradually darken the sclera. The process starts by using a narrow brush to go along the outlines of the sclera and the iris. This prevents the border of the eyes from being blurred. Additionally, this allows for faster and even coloration of the sclera. After that, a wider brush is used to darken the sclera repeatedly, three times in total. See figure 8 for the settings in Photoshop.

Figure 8

Settings for stimulus manipulation in Photoshop.



Note. A flow of 60% is used to fill the sclera. This is reduced to 10% to color the outline of the sclera and iris.

Eye Tracking & Software

A screen-based Tobii Pro Fusion was used to record eye movements at 120 Hz. It was attached to the bottom of a 25-inch 1920x1080 full HD monitor with a magnet strip which prevents unwanted movements. The eye tracker was calibrated using the Tobii pro eye tracker manager. The data recording, experimental setup and areas of interest (AOI) were done in Tobii pro lab. Open sesame was used to present the visual analog scale for likeability ratings. Data cleaning and analysis was done in Rstudio.

Measures

Most commercially available eye tracking systems process the raw data and present you with fixations and saccades. Especially Tobii pro lab offers a lot of preprocessed measurements. As we are specifically interested in the oscillation of the gaze between the eye region and other AOIs, the raw data is used as well as information about which AOI has been hit. Information on which AOI is hit is used to calculate the number of oscillations back and forth from the eye region.

Likeability Scores

Similar to the study of Mathur and Reichling (2016) a continuous visual analog scale (VAS) without graduation is used to determine likeability scores. A VAS consists of a 100-mm long horizontal line with verbal descriptions at each end. The verbal descriptions function as "word anchors" and express the extremes of a feeling. The question of the scale reads "To me, this face seems..." with the word anchors (less friendly, more unpleasant, creepy) on the left and (more friendly and pleasant, less creepy) on the right. The advantage of a VAS lies in them being perceived as a continuum with interval-scaled data. Respondents interpret equally sized intervals as two equally sized differences in affection. Compared to a Likert-type ordinal scale, this can provide more precise and psychometrically valid ratings (Reips & Funke, 2008). Participants will score each stimulus once after being exposed to it.

Procedure

On arrival, participants are welcomed and given a brief explanation of the following study. The true purpose of the study is hidden to prevent bias in the participants' responses. A generic explanation, comparing the likeability of humans and apes, was made up and is given instead. Following that, the participants were asked to read the consent form (see Appendix A), ask questions if they have any and sign the form. The Ethics Committee of the Faculty of Behavioural and Management and Social Sciences at the University of Twente approved this study. An oral explanation of the procedure of the experiment is given to the participant by the researcher. This was done as written instructions always resulted in questions or misunderstandings during the trial runs. The Participants were then asked to take a seat 60 cm away from a screen. Screen height is adjustable, to ensure similar comfort for all participants. Next, the eye tracker is calibrated in Tobii pro labs. There are two conditions with 40 stimuli each. Condition one contains the stimuli form the CFD and the generated primate pictures. Condition two contains the manipulated versions. Each stimulus is proceeded by a fixation point. The fixation point is placed in the center of the screen and displayed for 750 milliseconds. Each stimulus is presented for 2,5 seconds, even though similar UV effect curves have been reported for stimulus presentation times as low as 50ms (Slijkhuijs, 2017). This is done to allow for more eye tracking data to be collected. Limmer (2023) argued that a stimulus presentation time below 2 seconds can impact the data analysis negatively. After each stimulus, the likability response scale is presented and filled out by the participants. Participants are instructed to report their personal reaction to the stimuli. After the experiment's completion, participants are thanked for their contribution and given a short debriefing and explanation about its true goal.

Data Analysis

In order to make inferences on the scan path and oscillations different areas of the face must be distinguished from one another. Therefore, the Mouth, Nose and Eye region were separated into different Areas of Interest (AOI). The Tobii pro labs software was used to create a hand-drawn outline of the unaltered faces shown as stimuli in condition one. Afterwards, the three AOI's (Nose, Mouth, Eye region) were outlined by hand as well. The outline and AOI's were then copied and pasted onto the altered stimuli used in condition two. With this the scan path for each participant was established. The amount of oscillation between the eye region and the AOIs surrounding that area were calculated by using the raw data from Tobii pro labs that an AOI was hit at a certain point in time. The number of oscillations as well as the likeability scores were compared for both conditions in multi-level treatment effects models.

Likeability scores were tracked on continuous visual analog scale (VAS) without graduation. The stimulus presentation software Open sesame was used to create and administer the VAS during testing. A score ranging from 0 to 1 is assigned to the position of a respondent's cursor on the line.

Data had to be excluded if the eye-tracking recording failed due to miscalibration of a general malfunction of the software. Even if only one condition per participant is not correctly recorded, the participants data for both conditions is excluded from the study.

Results

Likeability scores

Figure 9 depicts a visualization of the group averages. A 2x2 multi-level treatment effects model was used to estimate the effects of species and congruency on likeability ratings. Table 3 depicts the population-level fixed effects estimates. Table 2 shows the average rating for each group of stimuli.

Table 2

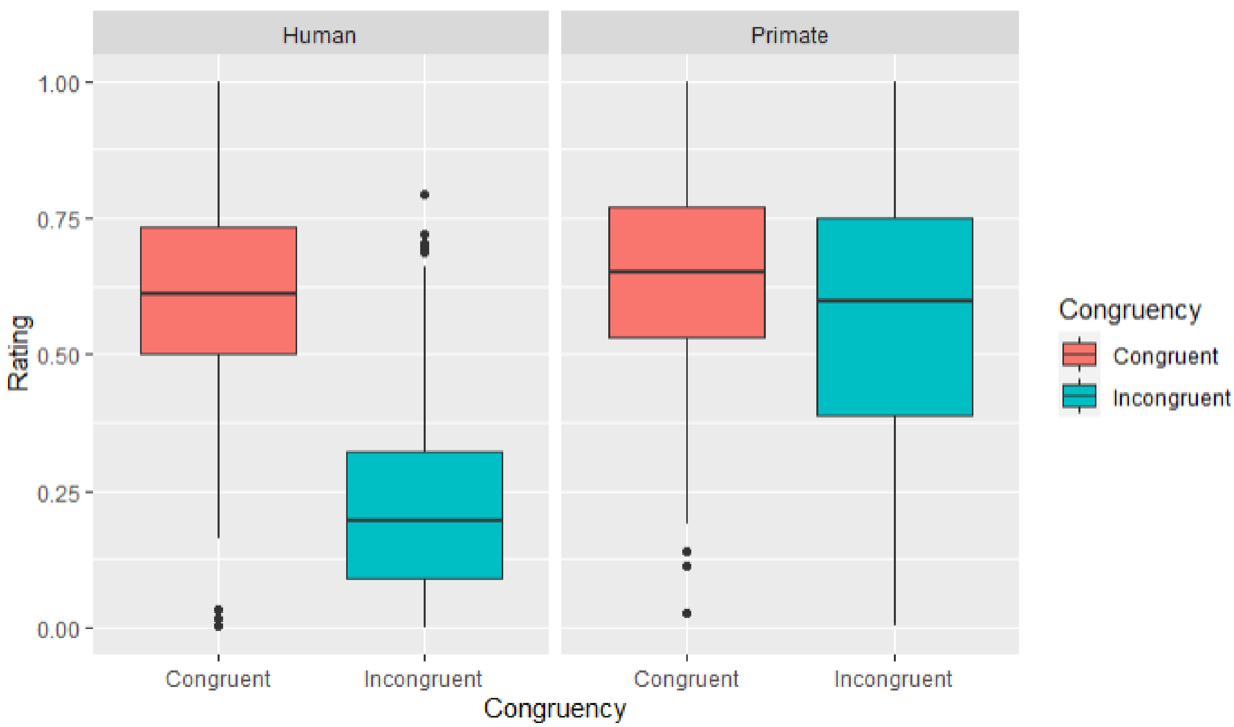
Average likeability scores for each group of stimuli.

Congruency	Species	Avg_rating
Congruent	Human	0.6086198
Congruent	Primate	0.6454410
Incongruent	Human	0.2192001
Incongruent	Primate	0.5755434

Note. Likeability ratings consist of a number between 1 and 0. 1 indicated the highest possible likeability rating, while 0 indicates the lowest.

Figure 9

Average likeability scale rating of human and primate stimuli across congruency conditions



Note. This visualization shows the average rating of stimuli belonging to the Human or Primate *group (non-human). Congruent stimuli, indicated by blue color, refer to faces with an eye-face mismatch.*

A first inspection of the group averages in Figure 9 and Table 2 indicates lower likeabilty ratings for incongruent human and primate stimuli. The ratings of human incongruent stimuli appear to be lower than those of incongruent primate stimuli.

Table 3

Multi-Level model for treatment effects of Species and Congruency on Likeability ratings, fixed effect estimates with 95% credibility limits.

fixef	center	lower	upper
Intercept	1.8405806	1.7490361	1.9363757
Congruency Incongruent	0.6764208	0.6281894	0.7305018
Species Primate	1.0370154	0.9617655	1.1215778
Congruency Incongruent : Species Primate	1.3789268	1.2534070	1.5028537

Note. Intercept refers to congruent human faces (not manipulated). Multiplicative effects are interpreted. 1 is the neutral element while a value <1 indicates a decrease and a value >1 indicates an increase.

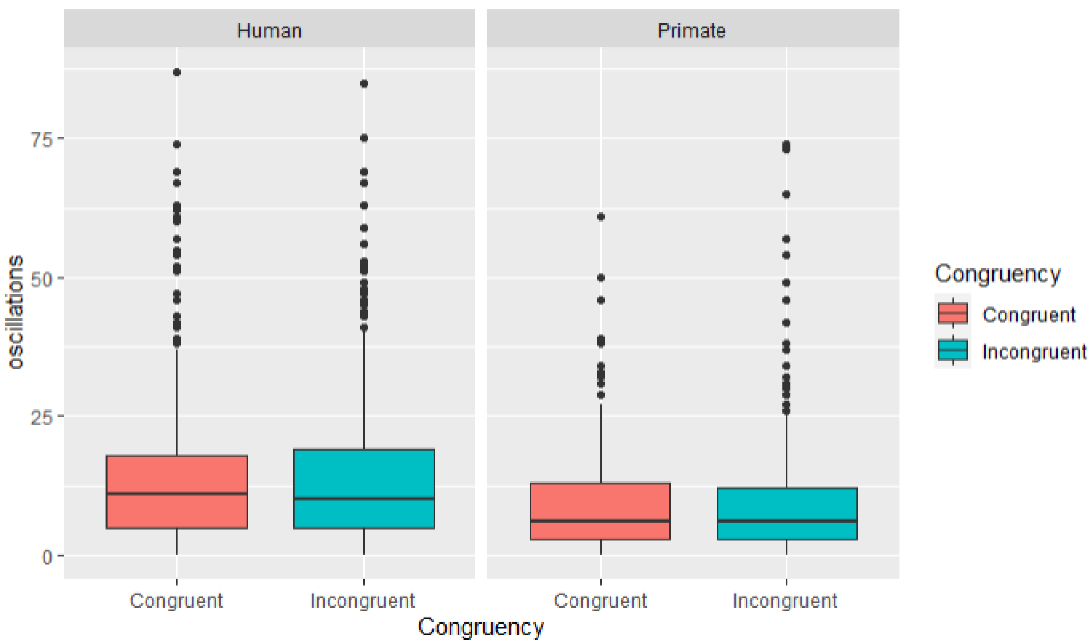
From the model (see Table 3) we can derive with sufficient certainty that incongruent human faces were rated lower than congruent human faces. Congruent primate faces were rated slightly higher than congruent human faces but not with sufficient certainty. The difference in likeability ratings for Incongruent and congruent primate faces is lower than for human faces. Incongruent primate faces were rated only slightly, but with sufficient certainty, lower than congruent primate faces. The manual manipulation of the human faces had a strong negative effect on likeability ratings while the AI generated human eyes on primate faces only had a small negative effect.

Oscillation count

The effect of eye region congruency and the species on the number of oscillations going from and to the eye region were estimated using a 2x2 multi-level treatment effects model. Table 4 displays the fixed effects estimates on a population level.

Figure 10

Mean of number of oscillations of human and primate stimuli across congruency conditions.



Note. This visualization shows the mean of number of oscillations on a given stimulus, Human or Primate group (non-human). Congruent stimuli, indicated by blue color, refer to faces with an eye-face mismatch.

An initial look at the mean average of number of oscillations in the current study reveals even amounts across congruent and incongruent stimuli (see Figure 10). Primate stimuli appear to invoke slightly less oscillations.

Table 4

Fixed effect estimates of a Multi-Level Model for the Effect of Species and Congruency on Number of Oscillations, with 95% credibility limits.

fixef	center	lower	upper
Intercept	12.7968498	10.6243237	15.436343
Species Primate	0.6386240	0.5526046	0.734306
Congruency Incongruent	0.9604461	0.8432188	1.090453
Species Primate : Congruency Incongruent	10.7634119	8.9589237	13.157383

Note. Intercept refers to Human faces with a congruent (not manipulated) eye region. Multiplicative effects are interpreted. 1 is the neutral element while a value <1 indicates a decrease and a value >1 indicates an increase.

According to the coefficients of this model (see Table 4), oscillations to the eye region occur, with sufficient certainty, less often when looking at primate faces with a congruent eye region compared to congruent human faces. The number of oscillations for incongruent human faces is only slightly lower, and with insufficient certainty, than for congruent human faces. The number of oscillations on incongruent primate faces does not differ with sufficient certainty from that of congruent primate faces.

Discussion

The current study investigates the role that a face-eye mismatch plays in triggering the UV effect. Our main hypothesis was that faces with an incongruent eye region (a face-eye mismatch) cause a feeling of uncanniness. Additionally, we investigate the variation in gaze behavior of participants when encountering stimuli that trigger the UV effect and stimuli that do not. The sub hypothesis of the current study was that incongruent stimuli trigger a cross-validation mechanism which results in increased oscillations to the eye region.

Likeability Rating

The likeability scores indicate which affective response the stimulus elicited in the participant. The first assumption was that congruent faces do not elicit a negative response, resulting in higher likeability ratings. The second assumption was that incongruent faces cause a negative, eerie, response and would result in lower ratings. The third assumption was that incongruent primate faces would cause a strong negative response, resulting in the lowest ratings.

The estimated effects of species and congruency on likeability in our multi-level model showed that congruent faces result in higher ratings than incongruent faces. This is in line with prior research and our assumptions. Therefore, the findings support our face-eye mismatch theory. It is important to mention here that, while the incongruent faces of both species were rated lower than their congruent counterparts, there is still a difference in the amount of impact the manipulation had for human and non-human stimuli. Incongruent human faces were rated lower than congruent human faces, the lowest of all conditions. This effect was less pronounced for non-human primate stimuli. Here, the difference in likeability scores between congruency conditions could not be shown with sufficient certainty. This finding contradicts earlier research by Stikker (2023). This study investigated the category confusion theory using animal, human and incongruent stimuli. The results showed that animal faces with human eyes were rated the lowest. Therefore, we would have expected our primate incongruent stimuli to receive the lowest rating. In an additonal condition, a pair of sunglasses was also used to cover the eyes of the stimuli. Stikker (2023) reported higher likeability estimates for animal stimuli wearing sunglasses than without. Therefore, Stikker argues that the eyes might not be the determining factor that triggers the UV effect. However, Stikker criticized the quality of the stimuli used, as some had low resolution and extreme facial expressions. She mentioned that facial expressions could lead to certain emotional responses that influence likeability scores even in the absence of a visible eye region. Due to this criticism the current study aimed to create a high-resolution stimulus set with neutral facial expressions. The neutral nature of our primate stimuli facial expressions does limit the effect they can have on likeability ratings.

Contrary to Stikker's findings, prior research by Geue (2021) supports the importance of the eye region in causing the uncanny valley effect. She investigated the UV effect in non-human primates of varying human likeness. This study was able to replicate the UV effect for primate faces and highlighted a shared characteristic in the eye region for stimuli that trigger the UV effect. Primate stimuli with a visibly white eyeball, the so-called sclera, seem to fall in the trough of the uncanny valley. A visibly white sclera is a characteristic of human eyes and is seldom seen in animals. The sclera is of darker color and less pronounced in non-human primates. This led Limmer (2023) to come up with the scleral-facial mismatch theory which can be seen as a derivative of the category confusion theory. Limmer argued that a scleral mismatch (white sclera in primate faces, dark sclera in human faces) triggers the UV effect. The results of his study are similar to that of the present study. Incongruent human faces were rated lower than their congruent counterparts. However, this effect was not found for primate faces. This contradicts his theory as lower likability scores were predicted for mismatched faces across human and non-human faces. Limmer explains these differences between human and non-human faces with unsuccessful experimental manipulation. A study by Kobayashi & Kohshima (2001) investigated the external eye morphology of primates and non-human primates. The results showed that non-human primate eyes differ from human primate eyes. Non-human primates have less visible sclera. Also, the area surrounding the eyes differs. For example, non-human primates have thicker brows, and the orientation of the eyes is less horizontal. Limmer argues that due to this smaller sclera, the manipulation would be less apparent in non-human stimuli. Thus, making the manipulation less effective. Limmer concluded that due to the scleral-facial mismatch not resulting in reduced likeability scores for non-humans, the scleral-facial mismatch theory was not fully supported by his results. However, Limmer states that the eye region appears to be the deciding factor for triggering the UV effect. The current study can be seen as a follow-up study to Limmer (2023) and attempted to improve the stimulus manipulation by using ai generation to let the eye regions of non-human stimuli appear more human like. Instead of small changes in the sclera the whole eye region was changed to create a more visible and impactful manipulation. Possibly due to these changes, a small but uncertain effect of the stimulus manipulation on likeability scores could be found for non-humans in our current study. The cause of feelings of uncanniness was hypothesized by Jentsch (1997) to be the inability of person to assign a known category to an observed entity. This is called the category confusion theory, in which our face-eye mismatch theory also has its roots. In terms of our current study that would mean, participants are struggling to identify the incongruent stimuli as human or primate. MacDorman and Chattopadhyay (2016) expanded on this idea for the use of artificially created stimuli. Their theory of realism inconsistency entails that feelings of uncanniness are caused by differing levels of humanlike realism in separate parts of one stimulus. In our current study that would mean the difference in realism between the manipulated eye region and the rest of the stimuli. If we now look at the way stimuli are manipulated in the current study, we can see that human faces are manually photographed and manipulated in Adobe Photoshop while the primate stimuli are generated and manipulated by AI in Adobe Firefly. In terms of realism, the difference between two ai generated parts is arguably lower than between a photograph and the manually darkened Sclera. Thus, it can be argued that the incongruent primate faces show lower realism inconsistency and therefore, elicit less feelings of uncanniness which results in more positive ratings. Additionally, MacDorman and Chattopadhyay (2016) reported a dislike of visibly artificial faces. Geue and Schmettow (2021) also showed a more positive response for biological faces with increasing similarity to humans. These findings would lead us to expect lower ratings for our non-human stimuli (AI generated) than for our human stimuli (photographs). However, that was not the case in the current study. Non-human primate faces, incongruent as well as congruent, received higher ratings than their human counterparts. This could indicate that the stimulus creation process was of such high quality that the resulting stimuli appear to be real photographs. However, three participants mentioned that the non-human stimuli appear artificial, hinting at the gloss of the fur in some instances. The findings of the current study are in line with Limmer’s results regarding the likeability of non-human faces. In both studies non-human faces were scored higher on the likeability scales than human faces. However, this is not in line with previous UV effect research as the most human-like stimuli usually get the highest likeability scores (MacDorman & Chattopadhyay, 2016). Stikker (2023) mentioned that non-human stimuli are sometimes perceived as funny. Some of the participants in the current study also made similar remarks. Participants described some primate stimuli as looking friendly or thoughtful, while some of the human stimuli were regarded as unfriendly or in a bad mood. In selecting human faces with a neutral expression, we attempted to diminish the effect of facial expression on the likeability ratings. This seems to have worked only partially as our non-human stimuli elicited more positive responses than predicted by prior research.

A possible explanation for the unexpectedly high ratings of the incongruent non-human primate stimuli could lie in the theory of mind. Humans can effectively share their interest in external stimuli and their underlying mental state with their gaze (H. Kobayashi and S. Kohshima, 2001; H. Kobayashi and S. Kohshima, 2008; Emery, 2000). Often, a look at the eyes of a person is sufficient to make inferences about their feelings or plans of action (Porta et al., 2012). When participants in the current study are labelling the facial expression on a non-human primate stimulus as thoughtful, they make inferences about the mental state of the depicted primate. Studies by Ghiglino et al. (2020) and Willemse et al. (2019) show that subtle eye movements and following gazes of humanoid robots can increase their perceived human likeness and their likeability. The reason for this could be that with increased information from the eye region more inferences on the internal mental model of the observed stimulus can be made. Subsequently, this shapes our perception of a perceived stimulus in a positive manner. If we apply this concept to the current study, our manipulation of primate stimuli consisted of generating human eyes onto the face while our human stimuli were manipulated by darkening the sclera. Even the inferences on the mental state of congruent primate stimuli, being labelled as thoughtful, could alter the perception of this stimulus in a positive way. Especially the incongruent primate stimuli, with their generated human eyes, would enable more inferences on their mental state as their incongruent human counterparts. Kobayashi and Kohshima (2008) argue that the dark shades of brown in the sclera of non-human primates serves two purposes. Firstly, reducing the reflected sunlight (glare) in their eyes, reducing the likelihood of being spotted by predators. Secondly, attempting to hide the gaze direction, reducing the likelihood of getting attacked by a predator. Therefore, a white sclera enables others to accurately discern the gaze direction (Kobayashi and Kohshima, 2008). This is, as discussed earlier, used in a social context to make inferences about mental states. We can argue that the dark sclera of incongruent human stimuli obstructs inferences on mental states, resulting in a more negative appraisal of the stimulus. On the other hand, the white sclera and human like eye region of the incongruent primate stimuli enables inferences on mental states based on gaze direction. This might lead to the perceived increased likeability for incongruent primate stimuli compared to incongruent human stimuli.

In conclusion, our findings support the face-eye mismatch theory. We were able to replicate Limmer's study with a new stimulus set and manipulate the eye region to cause a change in likeability ratings.

Oscillation Count

The main objective of using eye tracking measurements in this study was to determine whether there are differences in the gaze patterns on stimuli that either do or do not evoke feelings of uncanniness. Specifically, we are looking for possible differences of the number of oscillations to the eye region between normal human and primate faces, and their manipulated counterparts. Identifying such differences would enable us to use oscillation count as a proxy measure for likeability scores in UV research.

Our assumption was that looking at incongruent faces (Humans with dark sclera and primates with human eyes) results in an increased number of oscillations, back and forth from the eye region. Whether or not the feelings of uncanniness stem from category confusion or realism inconsistency, they cause the gaze to oscillate between parts of the face to check the assumptions made on first inspection. However, this assumption necessitates the use of higher cognitive processes to evaluate the stimulus.

We found that incongruent faces do not significantly increase the number of oscillations to the eye region. This contradicts our assumption. However, the number of oscillations was significantly lower for the primate than for the human stimuli. This is in line with prior research (Dupierrix et al., 2014; Emery, 2000). While those studies did not directly look at oscillations, they found a significant difference between the amount of visual attention that human and non-human primate eyes receive, with human eyes receiving far more attention. We could argue that the lower oscillation count on the eye region of primates in our study is a result of humans giving more attention to human eye regions in general, as they derive more meaning from them. Contrary to this line of thought would be the findings of Cheetham et al. (2013), who reported that for both human and non-human faces, incongruent faces would receive more fixations. More fixations would also require moving your gaze to the eye region from another place more often, which would in turn be considered as oscillations. However, our results did not show a significant difference in the number of oscillations between congruent and incongruent faces.

An explanation for the stable oscillation count across congruency categories could lie in the nature of the UV effect, namely whether fast or slow system processes cause the phenomenon. As mentioned earlier, our assumption about oscillations necessitates higher cognitive processes to occur. Humans would need to categorize a stimulus and then cross check it. Only if that validation process indicates a mismatch would a feeling of eeriness occur. However, research by Or and Wilson (2010), showed that humans can recognize a face in as short a time as 63ms. The current experiment used an exposure time of 2500ms to allow for higher cognitive processing to occur. Additionally, it was indicated by participants in the pilot test that 2000ms would be too short to "properly see the face". However, when increasing the exposure time to 2500ms, some participants wanted to rate stimuli before the exposure time had ended. This also indicates that 2500ms are more than sufficient to judge a face.

Face perception is regarded as a specialized process, which incorporates a holistic mechanism that differs from the way other objects are perceived (McKone, Kanwisher, & Duschaine, 2007). They found that people are able to assign a stimulus to a group (face, object) as soon as it is detected in their visual field. The participants were shown to be able to group objects under 20ms. Furthermore, a study by Bar, Neta, and Linz (2006) was conducted, which investigated how long humans need to perform a judgment of threat for faces with neutral expressions. A fairly consistent threat judgement could be made with stimulus exposure times as low as 39ms. In addition to these findings, Stone, Valentine, and Davis (2001) were able to show that stimuli can not only be categorized as objects or faces by humans under 20ms but also can be judged as good or evil. This shows that face perception is highly specialized and occurs very fast. If humans can judge a face to be good or evil in such a short time it is likely that they can also experience feelings of uncanniness due to those faces quickly. Thus, the UV effect is evoked by fast processing systems. Therefore, higher cognitive processes are likely not involved in the first judgment of a face and cannot prompt the observer to check their initial category assumption by comparing regions of the face.

In conclusion, our findings did not confirm our sub-hypothesis. We found that congruent and incongruent faces received a similar number of oscillations. Primate faces received less oscillations to the eye region than human faces. We thought of the eye region as a cross reference point, which is being used to clear category confusion. This would be marked by an increase in oscillations to the eye region when encountering incongruent faces. This would assume the use of higher cognitive processes. However, the number of oscillations was influenced by the species and not the congruency of the stimuli. These findings hint at fast system processes being the root cause of the UV effect. This is in line with previous research that indicates similar UV curvatures for exposure times between 2 seconds and as short as 50 milliseconds, also indicating the use of fast system processes (Slijkhuijs, 2017).

Strengths and Limitations

The current study was the first to investigate the number of oscillations as a possible supplementary measurement for eeriness responses. However, the sub-hypothesis was not supported by our findings, as there was no effect of the experimental manipulation on the number of oscillations. Therefore, using pupil dilation as an indicator for eeriness appears to be a more promising direction of research (Saneyoshi et al., 2022; Sharma & Vemuri, 2022). In contrast to the sub-hypothesis the results support the main hypothesis of the face-eye mismatch theory. The experimental manipulation of the improved stimulus set showed effects on the likeability ratings. The current study serves as a successful follow-up study to Limmer and offers further support for an evolutionary approach in the search for the UV effect's cause.

Limmer (2023) mentioned the presentation time of two seconds, being a potential cause for low fixation- and dwell-count values. This could cause the differences between conditions to be small or completely masked. We chose to increase the presentation time slightly by half a second as participants in the pilot run often regarded the presentation time as too short. However, the half second increase in presentation time led to some participants not looking at stimuli for the full duration and swapping to the rating scale prematurely. This resulted in less gaze data being collected despite longer presentation times. After inquiry by the researcher, they mentioned that they could appraise and rate stimuli before the presentation time ended.

Future Research

Regarding possible proxy measures for eeriness, future research should focus on other eye-tracking metrics than oscillation count. As mentioned earlier, pupil dilation seems to be a promising direction (Saneyoshi et al., 2022; Sharma & Vemuri, 2022).

With the successful AI based generation and manipulation of artificial stimuli new possibilities open up. The increase in stimulus resolution could benefit the search for relevant sub-areas of the eye region, like the sclera or brows. Additionally, the results of the current study support the evolutionary approach. Research on stimuli that vary in their evolutionary closeness to humans could make use of new high-quality stimuli. With even more advancement in the field of AI generation a follow up study to the current one could be conducted where the incongruent human stimuli are not only colored but also received primate like eye regions. It would be interesting to investigate the effect of a primate-like eye region in human stimuli on likeability ratings. It poses the question of whether it has the same detrimental effects as simple dark coloration of the sclera, or if we could observe a less negative impact like the human-like eye region for incongruent primate stimuli had.

Conclusion

The focus of the current study was to investigate the face-eye mismatch theory and the creation of a new stimulus set. The main hypothesis: "Does a mismatch in face appearance and eye region lead to a feeling of eeriness?" is supported by our findings. We found that faces with an incongruent eye region are perceived as less likeable than congruent faces. The effect of the eye region incongruency was more pronounced for human stimuli than for non-human primate stimuli. Overall, our findings suggest that our manipulation was successful in evoking the UV effect. However, the number of oscillations did not show an increase for incongruent faces. This leads us to reject the sub-hypothesis and eliminates oscillations as a possible proxy measurement for eeriness. Therefore, future research into the eye region and fast-system processes as cause for the UV effect and other proxy measures than oscillations are highly encouraged.

References

Bagepally, B. S. (2015). Gaze pattern on spontaneous human face perception: An eye tracker study. Journal of the Indian Academy of Applied Psychology, 41(Special Issue 3), 127-130.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. Emotion (Washington, D.C.), 6(2), 269–278. http://doi.org/10.1037/1528-3542.6.2.269

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another Advanced Test of Theory of Mind: Evidence from Very High Functioning Adults with Autism or Asperger Syndrome. Journal of Child Psychology and Psychiatry, 38(7), 813-822. https://doi.org/10.1111/j.1469-7610.1997.tb01599.x Baron-Cohen, S., Wheelwright, S.,

Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. Journal of Child Psychology and Psychiatry, 42(2), 241-251.

Beesley, Tom & Pearson, Daniel & Pelley, Mike. (2019). Eye Tracking as a Tool for Examining Cognitive Processes. 10.1016/B978-0-12-813092-6.00002-2.

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.

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., & Jancke, L. (2014). Perceptual discrimination difficulty and familiarity in the Uncanny Valley: more like a "Happy Valley". *Frontiers in psychology*, *5*, 1219. https://doi.org/10.3389/fpsyg.2014.01219

da Fonseca Grebot, I. B., Cintra, P. H. P., de Lima, E. F. F., de Castro, M. V., & de Moraes, R. (2022). Uncanny Valley Hypothesis and Hierarchy of Facial Features in the Human Likeness Continua: An Eye-Tracking Approach [Article]. Psychology and Neuroscience. https://doi.org/10.1037/pne0000281

Dupierrix, E., de Boisferon, A. H., Méary, D., Lee, K., Quinn, P. C., Di Giorgio, E., Simion, F., Tomonaga, M., & Pascalis, O. (2014). Preference for human eyes in human infants. Journal of Experimental Child Psychology, 123, 138-146. https://doi.org/https://doi.org/10.1016/j.jecp.2013.12.010

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. Neuroscience & Biobehavioral Reviews, 24(6), 581-604.

Geue, L. (2021). From Robots to Primates: Tracing the Uncanny Valley Effect to its Evolutionary Origin [Bachelor Thesis, University of Twente]. http://essay.utwente.nl/87564/

Ghiglino, D., Willemse, C., Tommaso, D., Bossi, F. & Wykowska, A. (2020). At first sight: robots' subtle eye movement parameters affect human attentional engagement, spontaneous attunement and perceived human-likeness. *Paladyn, Journal of Behavioral Robotics*, *11*(1), 31-39. https://doi.org/10.1515/pjbr-2020-0004

Grebot, Ivan & Cintra, Pedro & Lima, Emilly & Castro, Michella & de Moraes Jr, Rui. (2021). Uncanny valley hypothesis and hierarchy of facial features in the human likeness continua: An eye-tracking approach. 10.13140/RG.2.2.30364.21124.

Haeske, A. B. (2016). The Uncanny Valley: Involvement of Fast and Slow Evaluation Systems.

Ho C.-C., MacDorman K. F., & Pramono Z. A. (2008). *Human emotion and the uncanny valley: A GLM, MDS, and Isomap analysis of robot video ratings*. *In Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction, HRI '08* (pp. 169–176). Presented at the HRI '08, New York, NY. 10.1145/1349822.1349845

Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: the core of social cognition. Neurosci Biobehav Rev, 33(6), 843-863. https://doi.org/10.1016/j.neubiorev.2009.02.004

Itier, R. J., Villate, C., & Ryan, J. D. (2007). Eyes always attract attention but gaze orienting is task-dependent: Evidence from eye movement monitoring. Neuropsychologia, 45(5), 1019-1028.

Jentsch, E. (1997). On the psychology of the uncanny (1906). Angelaki, 2(1), 7-16. https://doi.org/10.1080/09697259708571910

Keeris, D., & Schmettow, M. (2016). Replicating the uncanny valley across conditions using Morphed and robotic faces. University of Twente.

Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. Journal of Human Evolution, 40(5), 419-435. https://doi.org/https://doi.org/10.1006/jhev.2001.0468

Kobayashi, H., & Kohshima, S. (2008). Evolution of the human eye as a device for communication. In Primate origins of human cognition and behavior (pp. 383-401). Springer.

Koopman, R. (2019). The Uncanny Valley as a universal experience: a replication study using multilevel modelling [Bachelor Thesis, University of Twente]. http://essay.utwente.nl/77172/

Limmer, L., & Schmettow, M. (2023). The scleral-facial Mismatch Hypothesis: The Role of a White Sclera in the Evolution of the Uncanny Valley Effect

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. http://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/https://doi.org/10.1016/j.chb.2008.12.026

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

Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. Cognition, 146, 22-32. https://doi.org/10.1016/j.cognition.2015.09.008

Matsuda, G., Ishiguro, H., & Hiraki, K. (2015). Infant discrimination of humanoid robots [Article]. Frontiers in Psychology, 6, Article 1397. https://doi.org/10.3389/fpsyg.2015.01397

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. Trends in Cognitive Sciences, 6(6), 255–260. https://doi.org/10.1016/s1364-6613(02)01903-4

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior research methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? Trends in Cognitive Sciences, 11(1), 8–15. http://doi.org/10.1016/j.tics.2006.11.002

Mori, M. (2012). The uncanny valley (K. F. MacDorman & N. Kageki, Trans.) IEEE Robotics and Automation, 19(2), 98-100. https://dx.doi.org/10.1109/MRA.02012.2192811

Moosa, M. M., & Ud-Dean, S. M. (2010). Danger avoidance: An evolutionary explanation of Uncanny valley. *Biological Theory*, *5*, 12-14.

Mori, M. (1970). The uncanny valley. Energy, 7(4), 98–100. http://doi.org/10.1109/MRA.2012.2192811

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

Porta, M. Ricotti, S., & Perez, C. J (2012). "Emotional e-learning through eye tracking," *Proceedings of the 2012 IEEE Global Engineering Education Conference (EDUCON)*, Marrakech, Morocco, 2012, pp. 1-6, doi: 10.1109/EDUCON.2012.6201145.

Reips, UD., Funke, F. Interval-level measurement with visual analogue scales in Internet-based research: VAS Generator. *Behavior Research Methods* **40**, 699–704 (2008). https://doi.org/10.3758/BRM.40.3.699

Saneyoshi, A., Okubo, M., Suzuki, H., Oyama, T., & Laeng, B. (2022). The other-race effect in the uncanny valley [Article]. International Journal of Human Computer Studies, 166, Article 102871. https://doi.org/10.1016/j.ijhcs.2022.102871

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. Psychological Science, 13(5), 402-409.

Siebert, R., Taubert, N., Spadacenta, S., Dicke, P. W., Giese, M. A., & Thier, P. A naturalistic dynamic monkey head avatar elicits species-typical reactions and overcomes the uncanny valley. Society for Neuroscience 2020, 7(4), 1-17. DOI: https://doi.org/10.1523/ENEURO.0524-19.2020

Sharma, M., & Vemuri, K. (2022). Accepting Human-like Avatars in Social and Professional Roles [Article]. ACM Transactions on Human-Robot Interaction, 11(3), Article 28. https://doi.org/10.1145/3526026

Slijkhuis, P. J. H. (2017). The uncanny valley phenomenon. A replication with short exposure times. [Master Thesis, University of Twente]. https://essay.utwente.nl/72507/1/Slijkhuis_MA_BMS.pdf

Stikker, I. M. M. (2023). The influence of eyes in the occurrence of the uncanny valley effect. [Bachelor Thesis, University of Twente]. http://essay.utwente.nl/95299/

Stone, A., Valentine, T., & Davis, R. (2001). Face recognition and emotional valence: processing without awareness by neurologically intact participants does not simulate covert recognition in prosopagnosia. Cognitive, Affective & Behavioral Neuroscience, 1(2), 183–91. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12467113

Willemse, C., Marchesi, S., & Wykowska, A. (2019). Robot Faces that Follow Gaze Facilitate Attentional Engagement and Increase Their Likeability. https://doi.org/10.3389/fpsyg.2018.00070

Wilson, V.A.D., Kade, C., Moeller, S., Treue, S., Kagan, I., Fischer, J., 2020. Macaque gaze responses to the primatar: a virtual macaque head for social cognition research. Front. Psychol. 11 (1645) https://doi.org/10.3389/fpsyg.2020.01645.

Appendix A

Informed Consent Form

Consent Form for the dark side of the manga effect

Please tick the appropriate boxes	**Yes**	**No**
Taking part in the study		
I have read and understood the study information dated 14.02.2024, or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	□	□

I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	□	□
I understand that taking part in the study involves the use of eye tracking devices.	□	□
Use of the information in the study		
I understand that information I provide will be used for data analysis in a master thesis and potential future research.	□	□
I understand that personal information collected about me that can identify me, such as [e.g. my name or where I live], will not be shared beyond the study team.	□	□
Future use and reuse of the information by others		
I agree that my information may be shared with other researchers for future research studies that may be similar to this study. The information shared with other researchers will not include any information that can directly identify me. Researchers will not contact me for additional permission to use this information. (Note: This separate consent is not necessary if you will only store and share deidentified data.)	□	□
I give the researchers permission to keep my contact information and to contact me for future research projects.	□	□

Signatures

_______________________ _______________________ _________

Name of participant [printed] Signature Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

__________________________ ___________________ _________

Researcher name [printed] Signature Date