The scleral-facial Mismatch Hypothesis:

The Role of a white Sclera in the Evolution of the Uncanny Valley Effect

Leonard Limmer

Master Thesis Human Factors & Engineering Psychology

April 2, 2023

First Supervisor: Dr. Martin Schmettow

Second Supervisor: MSc. Peter J. H. Slijkhuis

Faculty of Behavioural, Management and Social Sciences

University of Twente

Abstract

After its discovery in the field of robotics, the uncanny valley effect, the eeriness response to almost-but-not-quite human-like faces, has been demonstrated to apply to biological faces with increasing ancestral closeness to humans as well. The present study investigates the scleral-facial mismatch hypothesis, which suggests that the mismatch between human eyes (=with white sclerae) and a non-human face morphology cause the effect. To test it, 59 participants rated the likeability of a set of faces, each face was presented once with dark- and once with white sclerae, while gaze behaviour was recorded with eye tracking. Human faces with a scleral-facial mismatch were perceived as less likeable than congruent human faces. Compared to congruent faces, mismatched faces also received more visual attention towards the eye region. However, the pattern of reduced likeability for mismatched- compared to congruent faces was not found in non-human faces. Therefore, our findings partially support the scleral-facial mismatch hypothesis and are not sufficient to fully explain the evolutionary origin of the uncanny valley effect. We suspect that the eye region is indeed the responsible area, but not the sclera by itself.

Keywords: uncanny valley, evolutionary psychology, eye tracking

Table of Contents

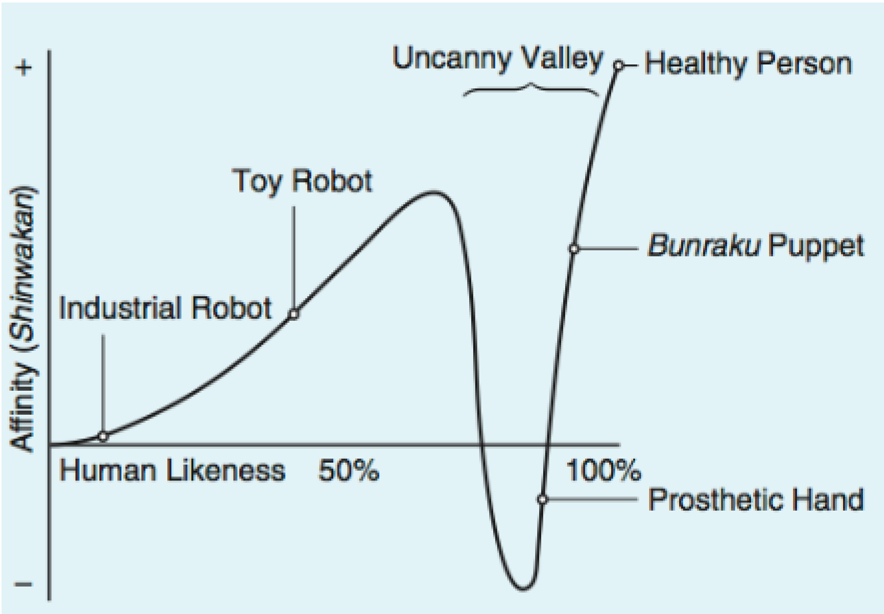
Introduction

The uncanny Valley Effect

The uncanny valley effect is the psychological response of eeriness to almost-but-not-quite human-like faces and was first hypothesised in 1970 by the Japanese roboticist Mori as a reaction to artificial (e.g., robot) faces. Mori proposed that when comparing entities with increasing degrees of human likeness, an observer's affinity would increase as well. When a point of almost-but-not-quite human likeness is reached, the observer's affinity significantly drops off. When plotted on a graph (see Figure 1), there is a clear valley in the curve, which together with the feelings of uncanniness provoked by entities falling into the valley, has given the effect its name. Beyond the valley, there is a sharp increase in affinity as human likeness reaches levels of being indistinguishable from human persons.

Figure 1

Original Uncanny Valley Effect Curve proposed by Mori (1970).



Note. This graph shows the proposed relationship between the human likeness of an entity and the associated affinity an observer would feel towards it. Taken from an English translation of Mori's Japanese essay.

Since Mori's initial description based on self-observation, the effect has been scientifically validated, both with robot faces and with computer animated faces (MacDorman & Chattopadhyay, 2016). Several possible explanations for the effect have been proposed, which can broadly be categorized into two lines of thought, depending on the cognitive system responsible: fast or slow (Haeske, 2016). Fast-system theories argue that the effect arises from evolved automatic, specialized and stimulus-driven processing occurring early in perception (MacDorman et al., 2009). This means that the process is sub-conscious, that it takes place without deliberate cognitive involvement or evaluation. Slow-system theories argue the opposite, that the effect is indeed based on conscious deliberation processes, which occur later in cognitive processing. This would make the effect the result of a "cognitive conflict" (Zhang et al., 2020).

More recent experimental data shows support for the fast-system explanation; Haeske (2016) was able to replicate the characteristic uncanny valley curve with stimulus presentation times of 100ms, which does not leave enough time for slow-acting processes to become involved. Additionally, the uncanny valley effect has been demonstrated to be universal (Koopman, 2019), which likely rules out the influence of socialization, which would be required for slow-acting processes to cause the effect. Because of this support for the fast-system branch of theories, the present paper also assumes a fast-system mechanism and the following overview on the state of research will focus on fast-system explanations. There are two hypothesized factors that explain how the uncanny valley effect could have evolved: threat avoidance and evolutionary aesthetics. The threat avoidance hypothesis builds on the fact that increased genetic similarity brings with it an increased risk of disease contraction (MacDorman et al., 2009). Therefore, the more human-like an entity is, the more sensitive humans are to any abnormalities it displays (Green et al., 2008), abnormalities that could indicate such diseases or genetic defects (MacDorman et al., 2009). Based on this sensitivity, the eeriness response could be a protective mechanism that inhibits humans from interacting with entities that are potentially dangerous, either immediately due to disease, or prospectively due to a lack of reproductive fitness. Reproductive fitness is also a central aspect of the hypothesised factor of evolutionary aesthetics. Facial attractiveness judgements (averageness, bilateral symmetry, skin quality) are, as a result of evolutionary selection pressure, used as indicators for health and fertility (Jones et al., 2004; Langlois & Roggman, 1990; MacDorman et al., 2009; MacDorman & Ishiguro, 2006; Rhodes et al., 1998). Inversely, a face lacking these indicators for fitness is perceived as unattractive and the associated negative emotional response serves to prevent mating with such an individual. Since eeriness is not only associated with fear, but also disgust (Ho et al., 2008), it is therefore hypothesised, that the eeriness response of the uncanny valley effect is the result of this instinct to produce healthy offspring to preserve one's genes.

After the focus of uncanny valley effect research had been on artificial faces for the longest part, Geue (2021) was able to demonstrate the effect with non-human Hominid (i.e., close relatives of humans) faces. This further supports an evolutionary approach to the explanation of the effect. In their study, they plotted a set of biological face stimuli (primates, hominids, humans) along ancestral closeness and found the typical uncanny valley curve replicated. Ancestral closeness was found to be congruent with human likeness, where the evolutionary closer a face was to human, the more likeable it was rated, analogue to how likeability increases with increasing human likeness. Just as previously demonstrated with artificial faces, this increase of likeability only went up until a point, where the faces of certain human ancestors and hominid relatives were perceived as uncanny and formed a valley in the likeability curve. For a detailed report, see Geue (2021). As a post-hoc observation, they noted that most of the faces that fell into the uncanny valley shared a characteristic that set them apart from stimuli before the valley: a white sclera. Therefore, the purpose of the present study is to gather further evidence into this notion. We want to find out if the eyes are indeed the determining facial feature for the uncanny valley effect.

The Role of the Eyes in social Cognition

The idea that the eyes might play a crucial role in causing the uncanny valley effect is tentatively supported by findings on the importance of the human eye in social cognition. The following section compiled from the review by Itier and Batty (2009) gives an overview on this importance.

Human eyes are the most important facial feature for all social non-verbal communication purposes, including the processing of emotions, establishing of identity, and the direction of attention of others (Itier & Batty, 2009). The inner features of the face, meaning the eyes, nose, and mouth, receive more visual attention during face perception than the outer ones, meaning the hair, face contour, forehead, and ears (Althoff & Cohen, 1999; Walker-Smith et al., 1977; Yarbus & Yarbus, 1967). Of these inner features, the eye region receives the most visual attention and is used to discern a variety of information, such as gaze, head orientation, identity, gender, facial expression, and age (Henderson et al., 2005; Itier et al., 2007; Janik et al., 1978; Laughery et al., 1971; Luria & Strauss, 2013; Philippe G Schyns et al., 2002). Additionally, every facial expression of a base emotion (sadness, happiness, fear, anger, surprise, disgust (Ekman & Friesen, 1978)) has an eye-region component that an observer relies on for recognition (Philippe G. Schyns et al., 2002). Even in the case of happiness, for which a smiling mouth is the primary indicator, a “fake smile” can only be recognized by cues in the eye region (a lack of characteristic wrinkles around the eyes). The importance of the eye region for recognizing emotions goes so far, that often, the isolated eye region is sufficient to recognize the basic emotions, but even more complex ones such as envy or guilt (Baron-Cohen et al., 1997; Baron-Cohen et al., 2001). The fact that such a variety of information can be gathered from the eye region and that this region subsequently is so central to social cognition might stem from the fact that this area shows the most variation between individuals. Both face detection and face recognition performance are impaired significantly by masking the eye region in comparison to masking any other facial feature (Lewis & Edmonds, 2003; McKelvie, 1976)

Since the human eye plays such a pivotal role in social cognition and the perspective of the present paper is one of evolutionary psychology, the next section is about the phenotypical configuration of the human eye and the potential evolutionary factors that made the human eye into the communicative asset it is today.

The Evolution of the human Sclera

Among extant primates, humans are the only species with a white sclera. The sclera of other primates are different shades of (dark) brown (Kobayashi & Kohshima, 2008). This unique trait warranted investigation, Kobayashi and Kohshima (2008) built the following argument: Since producing the pigment necessary for a dark sclera (which humans lack) costs energy, it must serves a purpose. On one hand, a dark sclera in non-human primates (and other mammals) can prevent “glare” from reflected sunlight that might otherwise give an animal’s position away. On the other hand, a dark sclera can obscure the precise direction of gaze (or at least making it harder to read), which the authors argue to be beneficial for deceiving predators. A prey animal’s chances of survival are higher if the predator thinks that the prey has its gaze on them. Conversely, the function of human’s white sclera is mainly social, according to the authors. More specifically, a white sclera surrounding a dark iris makes discerning the direction of gaze easier for the observer. This hypothesis is further supported by the fact that, among primates, the sclera in human eyes is most exposed. Moreover, out of 82 primate species investigated, homo sapiens is the only one with a colouration pattern in which the sclera colour is paler than the face- and iris colour, which results in the eyes being outlined against the face and the iris in turn being outlined against the rest of the eye. The authors refer to this as a “gaze signalling type” pattern (Kobayashi & Kohshima, 2008). By being able to read each other’s gazes, humans can communicate non-verbally, which enables or aids cooperative behaviour such as hunting in a group, while abnormal scleral coloration can also be an indicator of disease. The further importance of the eye region for human social behaviour having been outlined in the previous section. The authors reason that this gaze signalling structure of the human eye evolved, broadly, in two steps. The first step potentially came with human bipedalism and the associated shift from an arboreal habitat to the savanna. Because it increases the visual field horizontally, more precisely the visual field that could be accessed by moving the eyes without having to move the head, the eye became more horizontally exposed. Thereby, the sclera also became more exposed. This then laid the foundation for the second step, wherein the potential social benefits of a gaze signalling eye type as outlined previously put evolutionary pressure on extending this initially only slightly increased scleral visibility.

Present Study

This section shows how the results from social cognition and evolutionary anatomy might explain the uncanny valley effect. As seen in social cognition research, the human gaze is strongly drawn to the eyes, or the eye region in general. Humans judge a variety of characteristics of a face based on the eye region. Since the human eye looks unique among primates, it seems reasonable that by looking at the eye region, humans also judge species belongingness. More specifically, humans might recognize other humans by the eyes. The evolutionary benefit of being able to recognize a member of the own species at a glance is self-evident, for a variety of contexts such as threat avoidance, within-species rivalry, and mate selection, as mentioned above. At a time where the ancestors of modern humans shared their habitat with other hominids with similar looking eye regions, human's evolved ability to recognize their own kind by their unique eyes was no longer sufficient. It is unlikely that the eyes are the single feature by which this judgement of belonging to the same species is made, since humans have additional (facial) features that make them stand out from other primates and indeed from other hominids. Nonetheless, a fast-acting categorization mechanism based on a quick glance at the eye region would need a correction for false positives. We argue that the existing body of research suggests that the uncanny valley effect could be that corrective mechanism, the result of the brain receiving conflicting information regarding species categorization: human eyes in a non-human face. An alternative explanation would be that the sensitivity to within-species facial abnormalities also flags non-species members, meaning that any non-human hominid falling into the uncanny valley does so due to an overly active within-species threat recognition and avoidance mechanism, which would mean that these faces would be perceived as sick or otherwise “weird” or abnormal human, instead of as altogether non-human.

The purpose of the present study was to investigate the role that scleral coloration plays in the appraisal of biological faces. Specifically, we wanted to test whether a mismatch between scleral coloration and the rest of a face, i.e., human sclerae in a non-human face, causes the uncanny valley effect. We call this the scleral-facial mismatch hypothesis. In order to investigate this scleral-facial mismatch hypothesis, we created a new stimulus set based on the one used by Geue (2021) by adding a second version of each selected face, in which the scleral coloration is inverted. This process results in the following combination of facial features and expected effects on the observer:

Table 1

Scleral-facial Combinations and expected resulting Effects.

Facial Feature	Non-Human Face	Human Face
White Sclera	Uncanny Valley Effect?	Normal Human Face
Dark Sclera	Normal non-human Face	Does not occur in Nature – Uncanny Valley Effect?

Note. Table shows all four possible combinations of facial features, given overall face shape (human and non-human) and scleral colouring (white and dark), with the associated expected effect on the (human) observer. “Normal” in this case means unlikely to trigger an eeriness response.

The resulting stimulus set could then be presented to participants, and the effect of a manipulation of scleral colouration on the presence of an eeriness response could be measured. We expected that congruent faces, so non-human faces with dark sclerae and human faces with white sclerae, would not cause an eeriness response and would be rated as likeable. Mismatched faces, so non-human faces with white sclerae and human faces with dark sclerae, we expect to cause an eeriness response and be rated as less likeable.

Research Question and main Hypotheses

Does a scleral-facial mismatch cause the uncanny valley effect?

H1 “Congruent” faces, faces in which the scleral coloration matches the skull shape (human or non-human), do not cause an eeriness response.

H2 “Mismatched” faces, faces in which the scleral coloration does **not** match the skull shape (human or non-human), cause an eeriness response.

Additionally, we investigated participants’ gaze patterns, to see whether there were measurable differences between gaze behaviour for faces that cause the uncanny valley effect, versus faces that do not. This could serve to deepen the understanding of the effect’s mechanism, for example if the eye region would receive less visual attention in mismatched faces. Additionally, any identified unique gaze pattern could subsequently be used as a behavioural measure for the uncanny valley effect that does not rely on self-reported scales. A systematic literature review was conducted into existing uncanny valley effect research using eye tracking technology to inform the present study.

Eye Tracking Methodology in Uncanny Valley Effect Research: A systematic Literature Review

Since its initial description by Mori (1970) as a valley in the graph of the relation between the human likeness of an entity and the perceiver's affinity for it, the uncanny valley effect has been investigated in the context of artificial faces. Recently however, the evolutionary origin of the effect is being investigated, after Geue (2021) demonstrated that the effect also applies to a stimulus set of biological faces, where human likeness is equal to evolutionary closeness (on the range of primates – apes – hominids – humans). The next step in this line of research is to determine which facial features are responsible for triggering the effect. Eye tracking technology is an obvious choice for this inquiry. In order to inform that research, the present paper seeks to accumulate past uncanny valley effect research that made use of eye tracking and find out, which eye tracking variables have been used and why.

Literature Review: Methods

Definition of Terms

The uncanny valley effect (UCVE) is the psychological response of eeriness to almost-but-not-quite human-like faces. This reaction can be provoked by artificial faces such as robots or computer-generated images, as well as biological faces.

An area of interest (AOI) is a selected part or region of a stimulus, for which eye tracking metrics can be extracted from the raw data.

Research Questions

This review aims to inform how to measure the UCVE utilizing eye tracking. Specifically, the objective was to identify eye tracking measures that have already been used in UCVE research. Three research questions were derived from this objective.

(R1) Which eye tracking variables have been used to investigate the UCVE?

Answering R1 was the main goal of this review, to establish an overview of best practices and experiences with given methods, in order to inform subsequent research.

(R2) What was the theoretical background/reasoning for choosing a given variable?

R2 aims to investigate the feasibility of each discovered variable in light of more recent theoretical developments in UCVE research, specifically regarding the "quick vs. slow" debate.

(R3) How was each variable analysed?

R3 was included to gain further insight into the research process. Eye tracking can be used both for qualitative and quantitative analysis, so for a full understanding of the state of eye tracking research in the field of the UCVE, the analysis of each variable had to be included in this review.

Eligibility Criteria

Included were articles that were:

(i) Listed in the database Scopus

(ii) Written in English, peer reviewed and published no earlier than 2010.

(iii) Reported on an experimental study using eye tracking technology, investigating any aspect of the UCVE.

Excluded were articles that reported on experiments with non-human participants, were not available in full text, or did not utilize eye tracking as part of the experimental design.

Search Strategy

Scopus was searched using the string:

TITLE-ABS-KEY ("uncanny valley" AND "eye tracking")

Record Categorization Strategy

Not all terminology used in the discussed articles followed a unified system, but since that issue is not the focus of the present review, all concepts and measures mentioned in each article were translated into the scheme used in the present paper. For example, what is referred to as “ROI” for region of interest by Saneyoshi et al. (2022) is listed here as AOI for area of interest.

Literature Review: Results

The initial search on Scopus yielded 12 papers. Two of these papers were working with non-human primates and were therefore rejected. Three papers were not available in full text and the utilized eye tracking measures were not mentioned in the abstract, these papers were also rejected. Two papers used eye tracking technology for digitally animating human faces, but not in an experimental design investigating the UCVE, which is why these papers were also rejected. For an overview of the resulting included publications, see Table 2.

Table 2

Overview of publications included in the review, with title and year of publication, a summary of the content, and the used eye tracking metrics.

Title of Paper	Authors	Year of publication	Topic of Paper	Employed Eye Tracking Measures
The other-race effect in the uncanny valley	A. Saneyoshi, M. Okubo, H. Suzuki, T. Oyama, & B. Laeng	2022	The role of other-race bias in UCVE – Participants rated UCVE triggering stimuli faces of their own race as more unpleasant	Pupillary Diameter
Accepting Human-like Avatars in Social and Professional Roles	M. Sharma, & K. Vemuri	2022	Three experiments on digital avatar acceptance	pupil size variation, fixation count and pupil size by face, eye region, jaw and mouth, and torso AOI
Infant discrimination of humanoid robots	G. Matsuda, H. Ishiguro, & K. Hiraki	2015	Showing a robot, human, and android (in pairs) to infants and measuring looking behaviour	Gaze count by face, body, and “goal” AOI
Category processing and the human likeness dimension of the uncanny valley hypothesis: eye-tracking data	M. Cheetham, I. Pavlovic, N. Jordan, P. Suter, & L. Jancke	2013	Stepwise morphing between human and CGI faces to investigate changes in eye tracking data	Total number and dwell time of fixations to facial features
Uncanny Valley Hypothesis and Hierarchy of Facial Features in the Human Likeness Continua: An Eye-Tracking Approach	I. B. da Fonseca Grebot, P. H. P. Cintra, E. F. F. de Lima, M. V. de Castro, & R. de Moraes	2022	Based on previous study; Investigating hierarchical processing of facial features, depending on human likeness	Gaze dwell time on eye, nose, and mouth AOI

The following section addresses each one of the research questions based on the included publications.

(R1) What eye tracking variables have been used to investigate the UCVE?

Pupil dilation was measured in two out of the five papers, four papers used some form of AOI-based metric. AOI based metrics included gaze count, gaze dwell time, fixation count and fixation dwell time.

(R2) What was the theoretical background/reasoning for choosing a given variable?

Saneyoshi et al. (2022) measured pupil dilation as an "objective measure of affective response and surprise". They also asked participants to rate the pleasantness of stimuli on a scale as an additional UCVE measure, an operationalisation for which they cite Seyama and Nagayama (2007), who in turn link this back to Mori (1970), who originally described the UCVE. Sharma and Vemuri (2022) similarly cite the link of pupil dilation to "emotional response" as justification for using this measure in their first two experiments. In their third experiment, they recorded fixation count by AOI (face, eye region, jaw and mouth, torso) and also measured pupil dilation across these AOI, although there is no theory-based argument given for why they elected this specific eye tracking metric (over, for example, fixation time) and these specific AOI. M. Cheetham et al. (2013) used number of fixations and dwell time across eyes, nose, and mouth AOI to establish the hierarchy in which these features are processed for faces of varying human likeness. Their reason for picking this specific metric was to avoid data fishing by analysing all available eye tracking measures. Since the study by da Fonseca Grebot et al. (2022) further investigates and expands on the findings of M. Cheetham et al. (2013), they reported using the same metric. They note however, that fixation does not have defined, fixed parameters (a fixation being a cluster of gaze points over time X inside an area of Y degrees of visual angle with no available standardization for X and Y), which can lead to varying results when measuring fixation-based dwell time. Therefore, they use gaze dwell time per AOI (the proportion of gaze data, recorded at the known sample rate, of XY coordinates that fall within an AOI). Matsuda et al. (2015) used eye tracking for a preferential looking paradigm study design with infants, comparing a human, android, and robot stimulus. They divided mean gaze counts by the total gaze count for each pairing of these three stimuli to get a proportional “looking time” for the three AOI (face, body, “goal” = an object the person or robot was interacting with) for each stimulus. Their reasoning for using gaze count as a measure of time was that a gaze equals 3.3ms at their sampling frequency of 300Hz.

(R3) How was each variable analysed?

Saneyoshi et al. (2022) averaged the pupil diameter data from both eyes, excluded eye blink and saccade data and then computed the diametric change ratio for each stimulus compared to a baseline stimulus and then performed a 2 (participant culture) × 2 (stimulus face race) × 11 (stimulus eye size) mixed design analysis of variance. Sharma and Vemuri (2022) used a T-test ($p < 0.05$; 2-tailed) on the pupil size measures to compare artificial and human stimuli. In their experiment with AOI data, they created heatmaps from the fixation count data and calculated the average pupil size of all participants for each AOI of each stimulus, also after correcting with a baseline. Matsuda et al. (2015) applied the same AOI (face, body, “goal”) to each stimulus and performed statistical analysis for each pair of stimuli (human, android, robot). They conducted a three-way ANOVA with participant age group, stimulus, and AOI as factors, with the proportion of “looking time” as dependant variable. As mentioned above, da Fonseca Grebot et al. (2022) calculated gaze dwell time per AOI as the proportion of gaze data, recorded at the known sample rate, of XY coordinates that fall within an AOI, using gaze data from the entire screen. They then averaged dwell time per AOI for male and female faces, for each participant, for use as dependant variable in an ANOVA with AOI (eyes, nose, mouth), human likeness (artificial, boundary, human) and face gender (male and female face stimuli) as within-subject factors, as well as participant gender as only between-subject factor. M. Cheetham et al. (2013) discarded all blinks, as well as fixations that fell outside the face stimulus. They calculated total fixation number and cumulative total fixation duration for each AOI, for each stimulus, for each participant. Since the timeframes differed by stimulus and participant, they normalized the aforementioned values to "proportion of the total number of fixations" and "proportion of the total fixation duration" within each AOI. They then used the resulting values as dependent variables in respective two-way RM-ANOVA with human likeness (divided into three previously established levels) and AOI as factors, and gender as between-subject variable.

Literature Review: Discussion

As with any experimental design, the choice of variables depends on the questions the experiment is used to investigate. All studies reviewed here were investigating differences in the perception or reception of artificial and human stimuli and used eye tracking as an objective behavioural measure. The specific sub-question then informed the further narrowing down from available variables.

The UCVE was always mentioned, but not always central: two studies (Matsuda et al., 2015; Sharma & Vemuri, 2022) were comparing artificial and human stimuli in general, with no specific focus on the UCVE. Arguably, that could be reason enough to exclude them from this review, however, the UCVE was mentioned in these studies, as part of the theoretical background, because it is an inevitable aspect to consider when making these artificial vs. human comparisons, to explain the results or to inform decision making in stimulus selection. The methodology is therefore similar to the remaining studies, which specifically investigate the UCVE, and thus, they were included.

The use (or lack thereof) of AOI depended on what part of a stimulus was investigated. When merely comparing artificial to human stimuli in their entirety, changes in pupil dilation can indicate differences in emotional response. However, in most cases, authors were interested in specific features of a stimulus. Since a shared long-term goal in this line of research is to enable the making of better (i.e., less eery) artificial human-like entities, this is logical. For this purpose, AOI-based metrics were used. They allow for the isolation of specific parts of a stimulus and the subsequent analysis of any eye tracking metric in respect to that feature. What was included in a given AOI was, again, specific to the research question. Some studies investigated whole-body stimuli and would use broad AOI like the face and the torso. However, it seems to be agreed on that the face is the critical feature of an entity for causing the UCVE. Therefore, most studies used narrower AOI. It was common to separate the face into eyes, nose, and mouth regions. There seemed to be no clear-cut choice for which eye tracking metric to measure using these AOI based on theory. There was, however, the remark by da Fonseca Grebot et al. (2022) on fixation metrics having non-fixed parameters which could lead to differing results across studies. This is of course not specific to UCVE research, but could be generalized as advice to use metrics that require as little parametric post-processing as possible, i.e., to work with raw data as much as possible. On the other hand, it could be generalized as stressing the need for a well-documented data analysis process. As mentioned in their paper, there are a lot of available eye tracking metrics and including all of them in a given analysis without proper theoretical justification could be seen as data fishing. As for data analysis, there also seemed to be no clear consensus other than there being a mix of common statistical techniques for comparing conditions, such as T-tests, ANOVA, and regression analyses.

For future UCVE eye tracking research as described initially in the present paper, this review has the following implications: The investigation of specific facial features requires the use of AOI-based metrics. While it might be preferable to work with raw data as much as possible, there are a variety of metrics available that can be used, as long as they are used consistently across conditions and well-documented. For data analysis, all common statistical techniques appear to be viable, however, there was no precedence for the use of multi-level models.

Additional Research Question for present Study

Based on the results of the literature review on eye tracking methodology in uncanny valley research, it was possible to include the following secondary research question in the present study. Are there differences in gaze behaviour between faces that trigger the uncanny valley effect, and faces that do not? Since the present study investigated the role of specific features, eye tracking metrics had to be AOI based. This means the question could be narrowed down further. Are there differences in gaze behaviour towards the specific facial features of the eye region, the nose, and the mouth, between faces that cause- and do not cause the uncanny valley effect?

Methods

Participants

The present study was conducted between the 30th of September 2022 and the 21st of October 2022, with a convenience sample of 59 students from the University of Twente, 47 female and 12 male, between the ages of 18 and 25 (M = 20). Participants signed up through the university's test subject pool webservice (Sona Systems) and received participant credits for participation. The only eligibility requirement was unimpaired vision, since wearing glasses or contact lenses might interfere with eye tracking. Since the uncanny valley effect has already been demonstrated to be universal, sample characteristics were deemed irrelevant.

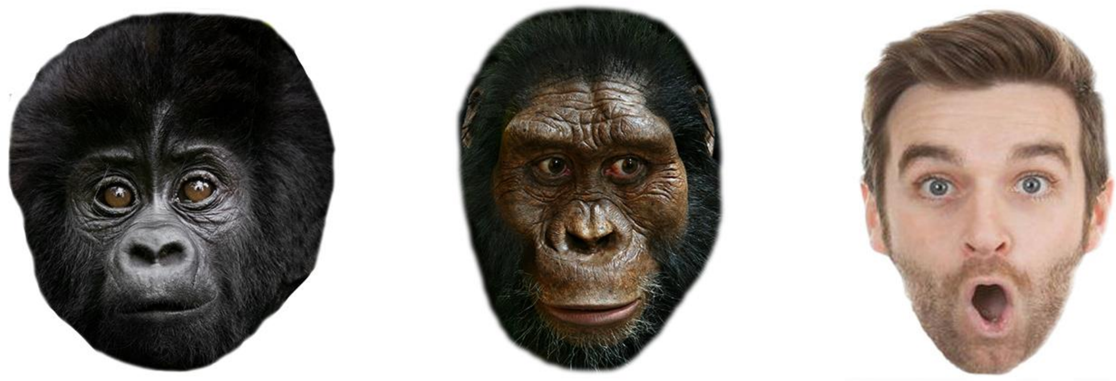
Materials

Stimuli

The stimuli set used is a selection of the set compiled by Geue (2021), which in turn was a compilation of primate faces from John Gurche's catalogue of hominid busts, the open access databases Global Biodiversity Information Facility and PrimFace, as well as free stock images of human faces showing different expressions. Each face was classified along its ancestral closeness to humans and has been rated for its human likeness (from 0 to 100) by four experts for the study by Geue (2021). The selection for the present study was done to narrow down the set to a more manageable size of 52 representative examples, since the occurrence of the uncanny valley effect with biological faces had already been demonstrated and did not require replication. Therefore, all robot faces were excluded. Next to the classifications made for the study by Geue (2021), all stimuli were classified along the colour of the sclera (dark or white), as well as their face morphology (whether the face read as human or non-human). The original labelling of the stimuli was kept the same, which is why their label numbers range from 1 to 100 despite only containing 53 stimuli. See Figure 2 for some examples of used stimuli.

Figure 2

Exemplary non-manipulated Stimuli from the present Study.



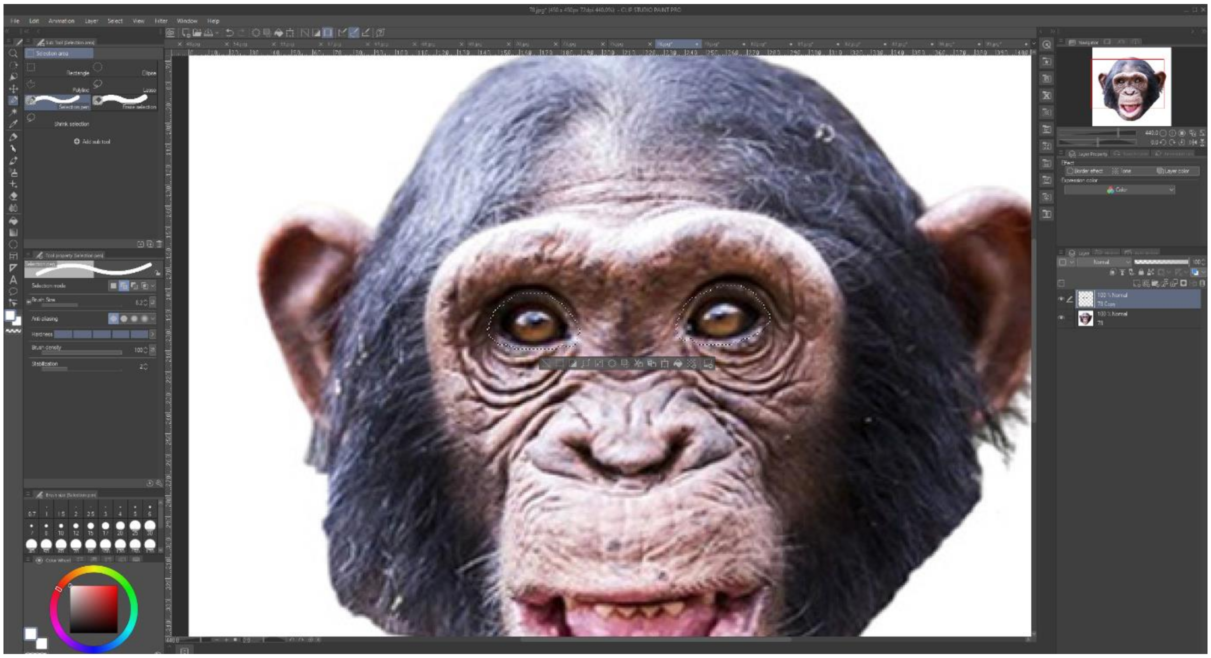
Note. These faces are ordered by ancestral closeness to humans (i.e., human likeness). They are, from left to right, a photograph of the face of a young gorilla, a reconstruction of an Australopithecus anamensis, and a photograph of a human face.

Manipulation

For each face, an altered version with the opposing scleral colouring was created. This was done using Clip Studio Paint. After consulting with a photography and image processing expert, it was determined that the alteration work would have to be done entirely manually, because the low resolution of the images did not allow for automation (such as using the wand selection tool, filters, stamping and the like). The images from the original stimulus set were used as reference. The following will detail the workflow that resulted as a combination of the aforementioned expert's advice and further testing and trial and error by the researcher. In order to work non-destructively, the eye region was copied to a new layer for processing, see Figure 3.

Figure 3

Stimulus Manipulation in CLIP STUDIO Paint, isolating and duplicating the Eye Region of a Stimulus.



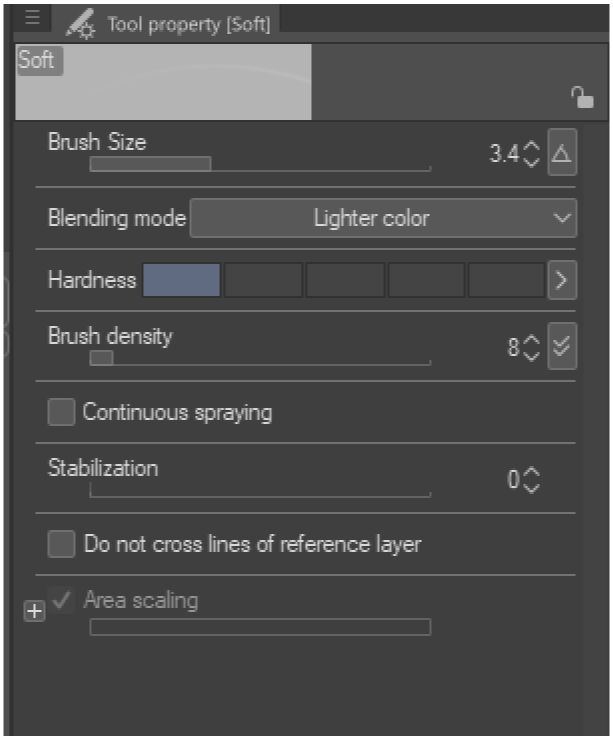
Note. This step was taken before further processing, to ensure non-destructive working.

For creating a dark from a white sclera, the airbrush tool was used at a low density and large tool size to gradually darken the sclera by iteratively applying a dark brown colour using either the “darken” or “darker colour” blending mode as suitable until the result was satisfactory. For creating a white from a dark sclera, the process was the same, with the blending mode being set to “lighten” or “lighter colour” and using white as the brush colour.

The airbrush tool was used with the settings shown in Figure 4.

Figure 4

Brush Settings in CLIP STUDIO Paint for creating a white from a dark Sclera.

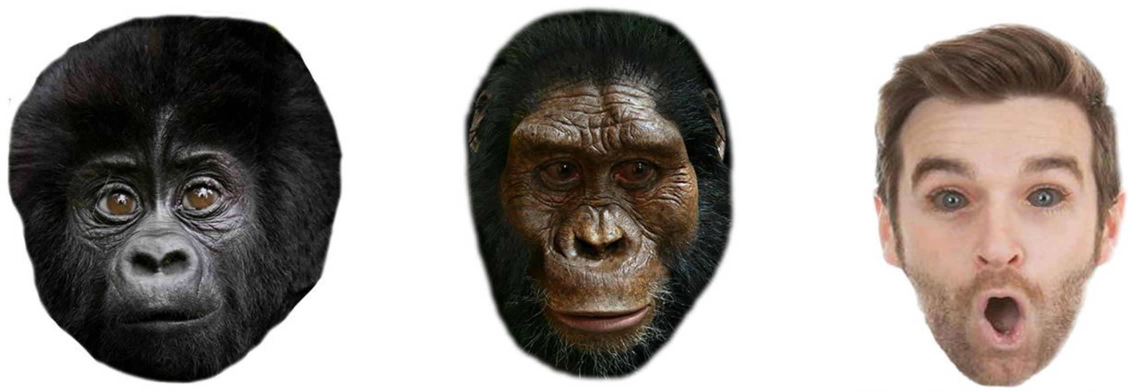


Note. The blending mode dictates how painting over an area takes into account the underlying layer. In this case, it was set to “lighter colour”, ensuring that details like light spots, shadows and reflections are not lost while the area is lightened. Blending mode “lighter colour” and “lighten” were used iteratively until the result satisfied.

The sclerae were then manually and iteratively drawn over, until it had the desired colour but still appearing spherical. This involved less whitening (or more darkening, respectively) in shadowed areas, not exceeding existing light reflections in brightness (or leaving those areas bright, respectively), and constantly zooming back out to ensure that the achieved effect fitted with the colour- and brightness pallet of the overall image. Because of the low resolution of the source images, some guesswork was involved in determining the border between the iris and the sclera. This introduced further difficulties, because in defining this border for both eyes, it was easy to make irregularity mistakes that would make the face appear cross-eyed. After adjusting the coloration to dark or white respectively, the edges of the sclera were cleaned up using a soft eraser tool. This gradually exposed the original image, specifically the iris and the area around the eyeball and any bright reflections, underneath the editing layer. To make this easier, the opacity of the editing layer was reduced, until the outlines of the eye's feature became visible as guidance from the layer below. This step allowed for the colour manipulation steps to be done less meticulously, saving time. For some examples of manipulated stimuli, see Figure 5.

Figure 5

Exemplary manipulated Stimuli from the present Study.



Note. These are the same images as in Figure 2, after their scleral coloration has been edited, making originally dark sclerae white and vice versa. The faces are, from left to right, an infant gorilla with white sclerae, a reconstruction of an Australopithecus anamensis with dark sclerae, and a human with dark sclerae.

Eye Tracking & Software

For recording eye movements, a screen-based TobiiPro Fusion recording at 120 Hz was used. It was attached to the bottom of a 27-inch 1920x1080 full HD monitor. Experimental setup, data recording and area of interest (AOI) drawing was done in iMotions 9.3. Further data treatment and analysis was done using RStudio.

Measures

Raw eye tracking data consists individual gaze points in space and time, at the system's recording frequency. However, most commercially available systems, such as the one in use for the present study, process that raw data into fixations and saccades. With the addition of defined AOI, the following variables could therefore be recorded:

- Number of fixations to each AOI,
- Number of dwells (dwell = sequence of fixations within the same AOI) per AOI
- AOI hit time (time in milliseconds after stimulus presentation at which an AOI was first regarded)

Additionally, a one-item likeability scale, adopted from Mathur and Reichling (2016), was used, asking participants to rate each face. The scale question reads "To me, this face seems…" with a response slider ranging from -100 (less friendly, more unpleasant, creepy) to 100 (more friendly and pleasant, less creepy). The tool for scale construction in iMotions re-coded this range into 0-15.

Procedure

Experiment

Participants were welcomed to the lab and provided with a brief explanation of the study's procedure. They were told that the purpose of the study was to measure emotional responses towards different faces because knowing the proper purpose of studying the role of the sclera in causing the uncanny valley effect could have biased their responses. Next, the participants were asked to read the consent form (see Appendix A), have any questions answered, and sign the form. Participants were seated at a desk ca 60cm in front of a screen. Screen height and tilt were adjustable, to ensure comfort and proper alignment. Once proper positioning was achieved, the eye tracker was calibrated, first within the TobiiPro device manager software, then in iMotions. For the experiment. Participants were shown 106 face stimuli, each for 2 seconds of exposure. There were two conditions with 53 stimuli each, condition 1 with the originals and condition 2 with the manipulated versions. All participants went through both conditions, in a randomized order. After each stimulus, a slide with the likeability response scale was presented. Participants were instructed to report their personal reaction to the stimuli and not for example the emotion the facial stimulus was displaying. After completing the experiment, participants were thanked for their participation and received a short debriefing about the purpose of studying the role of the sclera in the uncanny valley effect. The entire procedure from welcoming to seeing a participant off took ca. 30 minutes.

Data Analysis

AOI were created using iMotions built-in feature. For each stimulus, a polygonal AOI was drawn by hand for the entire face, and within that, the eye region (eyes and surrounding area including brows), the nose, and the mouth, and copied identically to the manipulated stimulus set, see Figure 6 for some examples.

Figure 6

Exemplary AOI from the present Study.

Note. Pictured are three examples of stimuli with overlayed AOI drawn in iMotions. There is one AOI each for the eye region ("ER"), the nose, and the mouth. The faces are, from left to right, that of a chimp, a reconstruction of an Australopithecus anamensis, and a human.

Eye tracking data recording did not work reliably in both conditions for all participants. In some cases, calibration repeatedly failed, and the experiment had to be conducted without recording eye tracking. In other cases, calibration was done successfully, but it turned out later that no eye tracking data had been recorded, for unknown reasons. Missing cases were excluded pairwise (i.e., from both conditions) per participant. Out of a total 20.796 recorded observations (where one observation was data from one AOI, on one stimulus, in one condition, from one participant), 2524 contained missing values and were removed, which was 12% of the total recorded data.

Likeability scale data was output by iMotions in a formatting suboptimal for further processing and analysis, insofar as putting every data point per participant in a single row. Therefore, scale data were atomized and then transformed such that every measurement (per participant, per stimulus, per condition) had a dedicated row in the resulting data frame.

Data analysis for each measure was done using multi-level treatment effects models.

Results

Likeability

The effects of face morphology and scleral coloration on likeability scale responses were estimated using a 2x2 multi-level treatment effects model, Table 3 gives the population-level (= fixed) effect estimates.

Table 3

Multi-Level Model for Treatment Effects of Scleral Coloration and Face Morphology on Likeability scale results, fixed effect estimates.

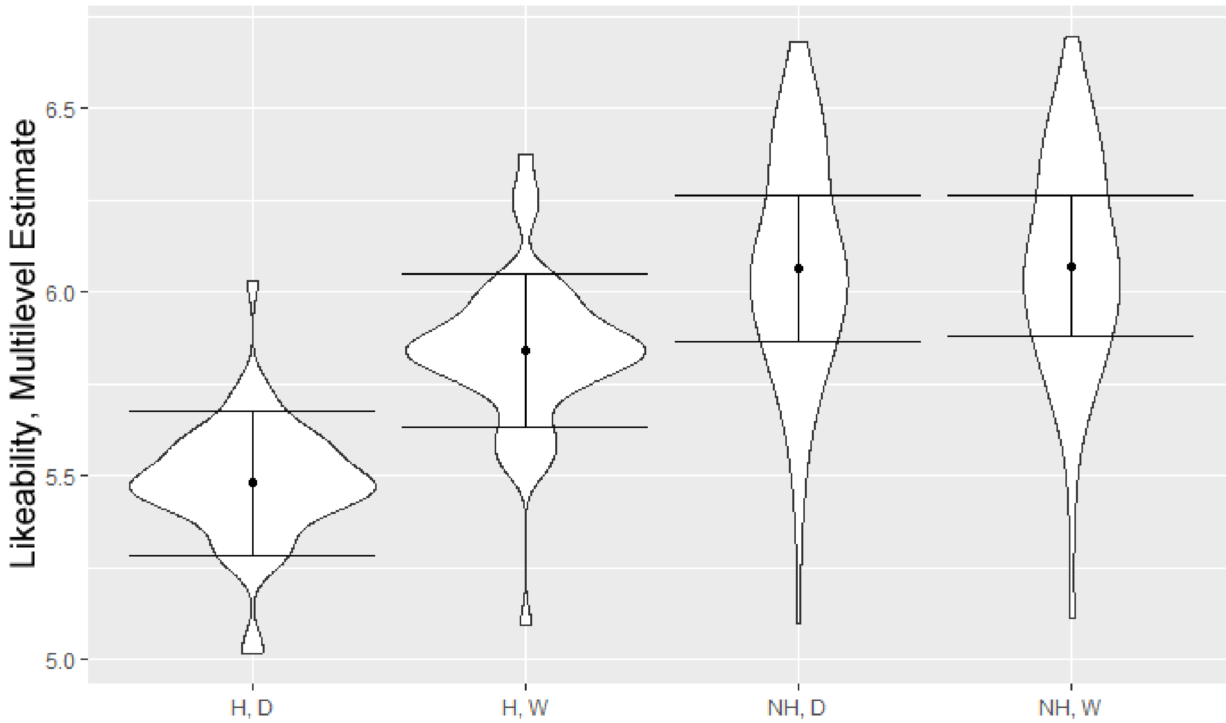
Parameter	Center Estimate	Lower Limit	Upper Limit
Intercept	6.0683169	5.8851781	6.2416643
White Sclera	0.0041441	0.1366185	0.1542915
Human Face	-0.5851138	-0.7679722	-0.3881733
White Sclera : Human Face	0.3536736	0.1164470	0.5924625

Note. Effect estimates are given with 95% credibility limits. Intercept refers to non-human faces with dark sclerae (no mismatch).

According to this model, faces with white sclerae were rated slightly, but significantly higher than (non-human) faces with dark sclerae. Human faces were rated significantly lower than non-human faces (with dark sclerae). The estimated effect of human faces with white sclerae (no mismatch) is significantly higher likeability than the estimates of the effect of the two factors in isolation. Absolute value estimates for each of the four face types were extracted from the model and are plotted in Figure 7.

Figure 7

Absolute Values from Multi-Level Model for Treatment Effects of Scleral Coloration and Face Morphology on Likeability scale results.



Note. This visualization shows the population-level (fixed) effect center estimates with 95% credibility limits, as well as the distribution of participant-level (random) effect center estimates. Face types are categorized by face morphology with H = human face, NH = non-human face, and scleral coloration with D = dark sclerae, W = white sclerae.

Separated into the four face types, these effect estimates show that human faces with either scleral coloration were perceived as less likeable. Between human faces, faces with dark sclerae, i.e., human faces with a scleral-facial mismatch, were rated as less likeable. The difference made by scleral coloration between non-human faces was comparatively small. Random effect estimates show a low spread for human faces, and a larger spread for non-human faces.

To check whether the manipulation of scleral coloration affected likeability ratings of all faces, not just human faces as demonstrated by the previous model (Table 4, Figure 7), manipulation was included as a factor in a 2x2x2 multi-level treatment effects model, Table 4 gives the population-level (= fixed) effect estimates.

Table 4

Multi-Level Model for Treatment Effects of Scleral Coloration, Face Morphology, and Manipulation on Likeability scale results.

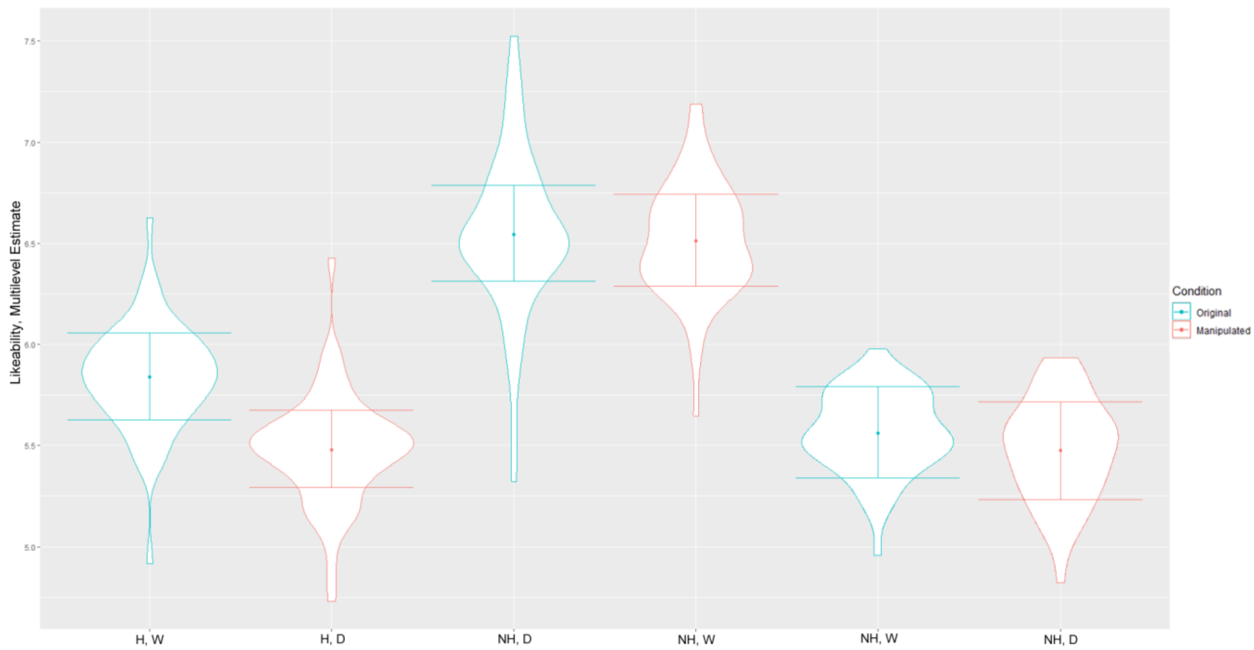
Parameter	Center Estimate	Lower Limit	Upper Limit
Intercept	6.5544712	6.3433907	6.7625296
White Sclera	-0.9820378	-1.1764678	-0.7915574
Human Face	0.0017539	-0.2222395	0.2139571
Manipulated	-1.0724509	-1.3011354	-0.8359231
White Sclera : Human Face	0.2724174	-0.0184277	0.5619883
White Sclera : Manipulated	2.0209935	1.7342936	2.3122416

Note. Effect estimates are given with 95% credibility limits. Intercept refers to original, non-human faces with dark sclerae (no mismatch).

Faces with white sclerae were rated significantly lower than faces with dark sclerae. Face morphology had no significant effect. Manipulated faces were rated significantly lower than original faces. There was no significant difference made by the factors white sclera and human face in conjunction (= congruent face) compared to individually. The factors white sclera and manipulation together had an estimated effect that caused significantly higher ratings than the individual effects of the two factors added up. Absolute value estimates for each of the four face types were computed and are plotted in Figure. They are additionally subdivided into original and manipulated versions, resulting in six cases (not eight, because there were no human faces with dark sclerae among the original stimuli). Absolute value estimates for each of the four face types were extracted from the model and are plotted in Figure 8.

Figure 8

Absolute Values from Multi-Level Model for Treatment Effects of Scleral Coloration, Face Morphology, and Manipulation on Likeability scale results.



Note. This visualization shows the population-level (fixed) effect center estimates with 95% credibility limits, as well as the distribution of participant-level (random) effect center estimates. Face types are categorized by face morphology with H = human face, NH = non-human face, and scleral coloration with D = dark sclerae, W = white sclerae.

When likeability scale ratings were separated into the four different face types present in the stimulus set, and it was accounted for whether a given face type was created by manipulation or present originally, the model showed that manipulating the scleral coloration had a greater effect on the likeability scale responses for human faces, compared to non-human faces. Manipulated versions of non-human stimuli were rated lower than their original versions, regardless of scleral coloration. Non-human faces with the same scleral coloration were rated significantly differently, depending on whether the scleral coloration was original or the result of manipulation. This explains why, under the previous model, the difference in likeability scale ratings depending on scleral coloration in non-human faces was so small. The spread of random effects observed for non-human faces was lower in this model, compared to the previous model. The spread of random effects for human faces was slightly greater than under the previous model.

Eye Tracking

Fixation Count

The effects of face morphology and scleral coloration on fixation count per AOI were estimated using a 2x2x4 multi-level treatment effects model, Table 5 gives the population-level (= fixed) effect estimates.

Table 5

Multi-Level Model for Treatment Effects of Scleral Coloration, Face Morphology, and AOI on Fixation Count.

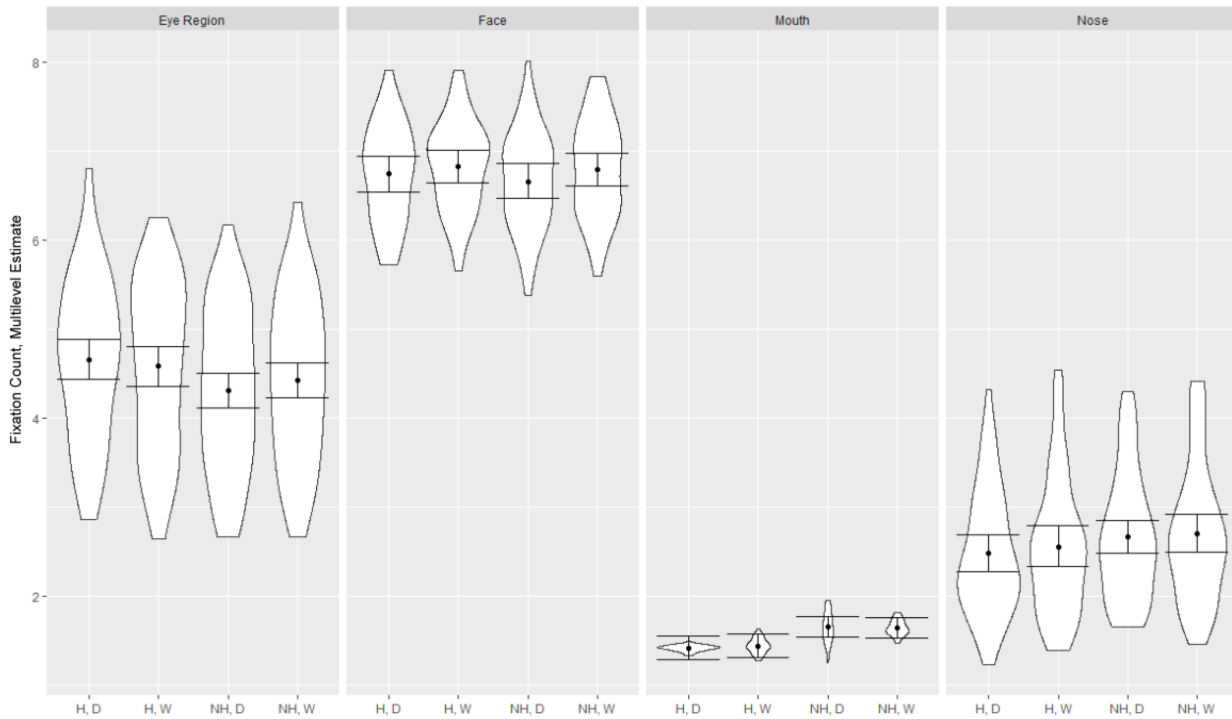
Parameter	**Center Estimate**	**Lower Limit**	**Upper Limit**
Intercept	4.3275735	4.1040953	4.5350387
White Sclera	0.1050644	0.0158669	0.1938630
Human Face	0.3417401	0.2265662	0.4583667
AOI Face	2.3474873	2.1399436	2.5545021
AOI Mouth	-2.6751448	-2.9164652	-2.4114658
AOI Nose	-1.6552553	-1.9255602	-1.3806991
White Sclera : Human Face	-0.1956439	-0.3824070	-0.0064480
White Sclera : AOI Face	0.0256093	-0.1031253	0.1536889
White Sclera : AOI Mouth	-0.1238744	-0.2604037	0.0188218
White Sclera : AOI Nose	-0.0618075	-0.1918181	0.0671896
Human Face : AOI Face	-0.2494542	-0.3889570	-0.1054596
Human Face : AOI Mouth	-0.6059041	-0.7714586	-0.4430727
Human Face : AOI Nose	-0.5294166	-0.6727507	-0.376888
White Sclera : Human Face :AOI Face	0.1381194	-0.0601093	0.3539961
White Sclera : Human Face : AOI Mouth	0.2464272	0.0028296	0.4916774
White Sclera : Human Face : AOI Nose	0.2358383	0.0047442	0.4824011

Note. Effect estimates are given with 95% credibility limits. Intercept refers to the eye region AOI on non-human faces with dark sclerae (no mismatch).

As seen in Table 5, (non-human) faces with white sclerae received slightly, but significantly more fixations to the eye region AOI than non-human faces with dark sclerae. Human faces also received significantly more fixations to the eye region AOI than non-human faces. The face AOI, which all other AOI were a part of, (necessarily) received the most fixations. The nose AOI received significantly fewer fixations than the eye region, the mouth AOI received significantly fewer fixations than the nose AOI. Human faces with white sclerae received slightly but significantly fewer fixations on the eye region AOI than the individual effect estimates for these two factors added up would predict. All human faces received significantly fewer fixations on the entire face AOI, the nose AOI, and the mouth AOI, than the individual effect estimates for each of the two factors added up, respectively, would predict. Human faces with white sclerae received significantly more fixations on the nose- and mouth AOI, than the individual effect estimates for each of the three factors added up, respectively, would predict. Absolute value estimates for each of the four face types, across the four AOI, were extracted from the model and are plotted in Figure 9.

Figure 9

Absolute Value Estimates from Multi-Level Model for Treatment Effects of Scleral Coloration, Face Morphology, and AOI on Fixation Count.



Note. This visualization shows the population-level (fixed) effect center estimates with 95% credibility limits, as well as the distribution of participant-level (random) effect center estimates. Face types are categorized by face morphology with H = human face, NH = non-human face, and scleral coloration with D = dark sclerae, W = white sclerae.

As can be seen in Figure 9, the overall hierarchy of fixation count was consistent across face types: the entire face (necessarily) received the most fixations. Of the facial features, the eye region received the most fixations, even though this was subject to large individual differences (random effects). Human faces with dark sclerae (mismatch) received the most fixations to the eye region AOI. Similarly, between non-human faces, those with white sclerae (mismatch) received more fixations on the eye region AOI than those with dark sclerae (congruent). For the entire face, both between human- and non-human faces, mismatched faces received less fixations. Individual differences here were still large, but less so than on the eye region AOI. Non-human faces received more fixations towards the mouth AOI. Individual differences were small and showed the lowest spread of all AOI. Non-human faces also received more fixations towards the nose AOI. For both human-and non-human faces, those with white sclerae received more fixations than those with dark sclerae. Individual differences were larger than for the face AOI, but smaller than for the eye region AOI.

Dwell Count

The effects of face morphology and scleral coloration on dwell count per AOI were estimated using a 2x2x4 multi-level treatment effects model, Table 6 gives the population-level (= fixed) effect estimates.

Table 6

Multi-Level Model for Treatment Effects of Scleral Coloration, Face Morphology, and AOI on Dwell Count.

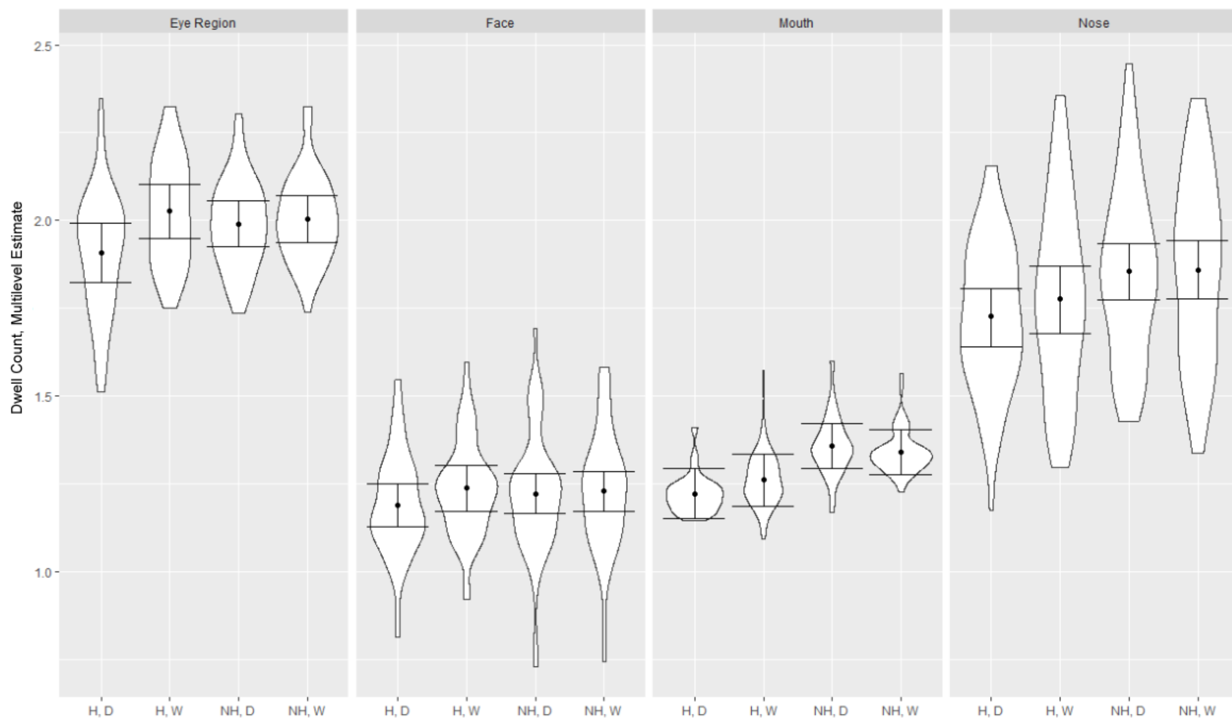
Parameter	Center Estimate	Lower Limit	Upper Limit
Intercept	1.9904882	1.9327690	2.0487906
White Sclera	0.0154546	-0.0283521	0.0566965
Human Face	-0.0777497	-0.1307182	-0.0223419
AOI Face	-0.7669965	-0.8338263	-0.6999832
AOI Mouth	-0.6312289	-0.6979258	-0.5609178
AOI Nose	-0.1333965	-0.2139084	-0.0465783
White Sclera : Human Face	0.0975805	0.0236251	0.1705891
White Sclera : AOI Face	-0.0044388	-0.0642558	0.0528070
White Sclera : AOI Mouth	-0.0315383	-0.0943199	0.0336881
White Sclera : AOI Nose	-0.0152687	-0.0741081	0.0465972
Human Face : AOI Face	0.0489577	-0.0169735	0.1160342
Human Face : AOI Mouth	-0.0748751	-0.1491556	0.0002062
Human Face : AOI Nose	-0.0532832	-0.1232021	0.0141283
White Sclera : Human Face :AOI Face	-0.0635306	-0.1566865	0.0342675
White Sclera : Human Face : AOI Mouth	-0.0295286	-0.1391634	0.0777208
White Sclera : Human Face : AOI Nose	-0.0430154	-0.1454118	0.061258

Note. Effect estimates are given with 95% credibility limits. Intercept refers to the eye region AOI on non-human faces with dark sclerae (no mismatch).

Human faces received less dwells to the eye region AOI than non-human faces. The eye region received significantly more dwells than the other AOI. The entire face was, on average, dwelled on only once. Human faces with white sclerae received less dwells than the individual effect estimates for these two factors added up would predict. There were no other significant effects. Absolute value estimates for each of the four face types, across the four AOI, were extracted from the model and are plotted in Figure 10.

Figure 10

Absolute Value Estimates from Multi-Level Model for Treatment Effects of Scleral Coloration, Face Morphology, and AOI on Dwell Count.



Note. This visualization shows the population-level (fixed) effect center estimates with 95% credibility limits, as well as the distribution of participant-level (random) effect center estimates. Face types are categorized by face morphology with H = human face, NH = non-human face, and scleral coloration with D = dark sclerae, W = white sclerae.

Figure 10 shows that human faces with dark sclerae (mismatch) received less dwells than all other face types, across all AOI. Non-human faces received more dwells on the mouth- and nose AOI than human faces. The spread of participant-level deviations (random effects) was very large on the nose AOI, less so on the eye region AOI and face AOI, and small on the mouth AOI.

Hit Time

The effects of face morphology and scleral coloration on hit time per AOI were estimated using a 2x2x4 multi-level treatment effects model, Table 7 gives the population-level (= fixed) effect estimates.

Table 7

Multi-Level Model for Treatment Effects of Scleral Coloration, Face Morphology, and AOI on Hit Time in milliseconds after stimulus presentation.

Parameter	Center Estimate	Lower Limit	Upper Limit
Intercept	230.86388	204.054317	256.3034097
White Sclera	-19.15696	-38.755783	2.4644293
Human Face	-26.35079	-52.688156	0.6505304
AOI Face	-173.29391	-197.765520	-147.1491074
AOI Mouth	515.17451	455.393242	571.4603095
AOI Nose	149.29575	100.230159	199.1858247
White Sclera : Human Face	24.55416	-12.753792	60.2225394
White Sclera : AOI Face	13.20098	-17.503463	40.8577770
White Sclera : AOI Mouth	44.67630	7.118863	83.1446876
White Sclera : AOI Nose	52.19688	18.517374	85.1266354
Human Face : AOI Face	22.01564	-11.717138	54.6654611
Human Face : AOI Mouth	217.09897	168.525655	265.7733064
Human Face : AOI Nose	57.89071	13.297067	101.7528878
White Sclera : Human Face :AOI Face	-15.08760	-62.693671	31.9288947
White Sclera : Human Face : AOI Mouth	-129.54211	-188.949267	-71.2618454
White Sclera : Human Face : AOI Nose	-16.38058	-78.116615	44.0630101

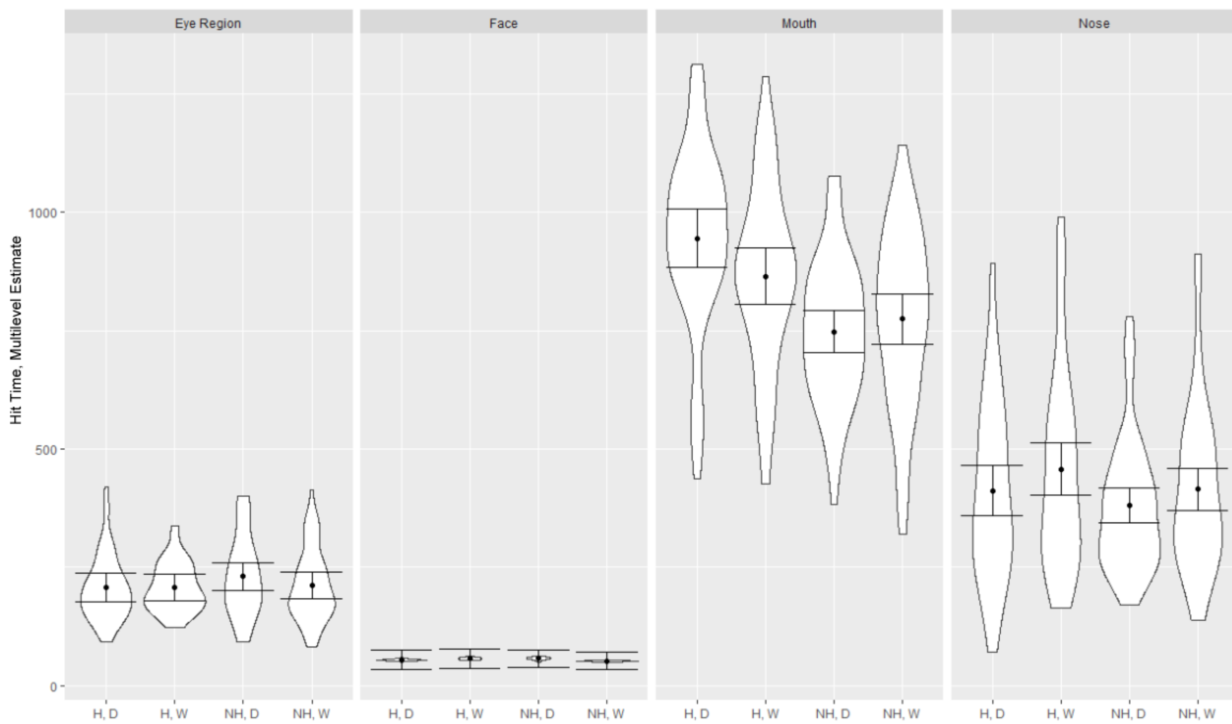
Note. Effect estimates are given with 95% credibility limits. Intercept refers to the eye region AOI on non-human faces with dark sclerae (no mismatch).

The face AOI was hit significantly faster than all other AOI. Between the facial features, there was a consistently significant order in which AOI were hit. The eye region AOI was hit first, the nose AOI significantly later, and the mouth AOI significantly later yet. The mouth- and nose AOI of (non-human) faces with white sclera were hit significantly later than the individual effect estimates for these two factors added up, respectively, would predict. The mouth- and nose

AOI of human faces (with dark sclerae) were hit significantly later than the individual effect estimates for these two factors added up, respectively, would predict. The mouth AOI of human faces with white sclerae was hit significantly earlier than the individual effect estimates for these three factors added up, respectively, would predict. Absolute value estimates for each of the four face types, across the four AOI, were extracted from the model and are plotted in Figure 11.

Figure 11

Absolute Value Estimates from Multi-Level Model for Treatment Effects of Scleral Coloration, Face Morphology, and AOI on Hit Time in milliseconds after Stimulus Presentation.



Note. This visualization shows the population-level (fixed) effect center estimates with 95% credibility limits, as well as the distribution of participant-level (random) effect center estimates. Face types are categorized by face morphology with H = human face, NH = non-human face, and scleral coloration with D = dark sclerae, W = white sclerae.

Figure 11 shows that, on the population level, the face AOI was always hit first, then the eye region-, then the nose-, then the mouth AOI. The eye region AOI of non-human faces with dark sclerae was hit later than that of other face types. Individual differences were visible, but small. The face AOI was hit consistently fast, with no noticeable individual differences. The mouth AOI of non-human faces was hit faster than that of human faces. Between human faces, the mouth AOI of faces with white sclerae was hit faster than that of faces with dark sclerae. For non-human faces, the inverse was the case. Individual differences showed a large spread. The Nose AOI was, for both human- and non-human faces, hit faster on faces with dark sclerae. Individual differences here show the largest spread of all AOI, such that some individuals apparently hit the nose AOI first on human faces with dark sclerae.

Discussion

The purpose of the present study was to investigate the role of scleral coloration in causing the uncanny valley effect. Our main hypothesis was that human-like eyes (=with white sclerae) in a non-human face cause the effect, thus naming it the scleral-facial mismatch hypothesis.

Likeability Scale Results

The first sub-hypothesis was that congruent faces do not cause an eeriness response. The second sub-hypothesis was that mismatched faces cause an eeriness response. Based on these two hypotheses, we expected high likeability ratings for congruent faces, and low likeability ratings for mismatched faces. Since Geue (2021) found significant individual differences in scale anchoring, the results from the likeability scale responses will be discussed in relation to one another, not in relation to their absolute positions on the scale. Therefore, summarizing the two sub-hypotheses, we expected higher likeability ratings for congruent faces, than for mismatched faces.

The multi-level model that estimated the effects of the factors of face morphology and scleral coloration showed that mismatched human faces (so, with dark sclerae) were rated as significantly less likeable than all other face types. This supports the scleral-facial mismatch hypothesis, which can be understood as a derivation of the category confusion hypothesis. Originally proposed by Jentsch (1997), this hypothesis states that feelings of uncanniness result from an observer's inability to clearly assign an entity to a known category. MacDorman and Chattopadhyay (2016) fount this insufficient to explain the uncanny valley effect in artificial stimuli and instead proposed that realism inconsistency within a stimulus explains the effect. Realism inconsistency meaning when different parts of the same stimulus differ in their respective level of humanlike realism. Applied to the present study and the case of biological faces, this can be extrapolated to the scleral-facial mismatch, where different parts of the stimulus, i.e., the scleral coloration and the rest of the face, differ in their respective levels of human likeness. This can be further specified. Existing findings from face processing research state that humans process faces primarily holistically. In other words, the overall "gestalt" of a face is always processed first, only then are individual facial features analysed (see for example Maurer et al., 2002). So, if the overall face shape is categorized as human, and then a facial feature, the coloration of the sclera, is subsequently categorized as non-human, category confusion ensues and manifests in form of an eeriness response. Our results are thereby in line with existing research, as well as marking the eye region as the responsible facial feature for the effect in biological faces. This, in turn, fits in with previous findings from social cognition research, which has repeatedly found the eye region to be the most important facial feature for a variety of classification tasks (Henderson et al., 2005; Itier et al., 2007; Janik et al., 1978; Laughery et al., 1971; Luria & Strauss, 2013; Philippe G Schyns et al., 2002). The same effect of a scleral-facial mismatch was, however, not found for non-human faces. This contradicts the scleral-facial mismatch hypothesis, which predicted lower likeability for all mismatched faces, including non-human faces with white sclerae. We found that this was due to an, in this respect, unsuccessful experimental manipulation. Where in human faces, introducing a scleral-facial mismatch via manipulation did indeed lead to lower likeability ratings, manipulating the scleral coloration of non-human faces had an inconsistent impact on likeability ratings. While non-human faces were rated lower when the sclera was manipulated white (so, when a mismatch was introduced), likeability ratings for non-human faces that originally had white sclerae did not improve when the mismatch was resolved by manipulation. There was a significant difference in likeability ratings between the original versions of non-human faces, depending on scleral coloration. Among original non-human faces, ones with a scleral-facial mismatch (so, white sclerae) were rated significantly lower in likeability than congruent faces (so, ones with dark sclerae). This would support the scleral-facial mismatch hypothesis, but the experimental manipulation was not successful in isolating scleral coloration as the cause of this difference in likeability ratings. For non-human faces with originally dark sclerae, the small difference made by manipulating scleral coloration (compared to the effect it had in human faces) might be because they lack the uniquely exposed sclera of human (and human-like) faces ("gaze signalling type", Kobayashi & Kohshima, 2001). That might have made it less obvious whether the sclera was white or dark. I.e., in non-human faces, the manipulation might have been less obvious and its effect therefore smaller than in human faces. This, however, does not account for the lack of impact on likeability ratings made by manipulating the scleral coloration in non-human faces with originally white sclerae, because their eye morphology is of the “gaze signalling type”. For these faces, the scleral coloration (and therefore, the manipulation) must have been as noticeable as for human faces. Therefore, the lacking effect of the manipulation must have had other reasons. As noted by Kobayashi and Kohshima (2001), not just the scleral coloration, but the entire eye morphology is different between humans and non-human primates. Therefore, manipulating scleral coloration in non-human faces with originally white sclerae might have been unsuccessful in masking the eyes’ overall “humanness”, thus not completely resolving the mismatch between the eye region and the face. On the contrary, it might have introduced a second mismatch, between eye morphology and scleral coloration, in addition to the existing mismatch of the human-like eye region in a non-human face. In summary, the manipulation was (apparently) successful in introducing a scleral-facial mismatch in human faces, which yielded the expected change in likeability scale ratings, but unsuccessful in introducing or resolving a scleral-facial mismatch in non-human faces, indicated by a lack of change in likeability ratings. This makes the present study unfit to explain the differences in likeability ratings between (original) face types found in the work of Geue (2021). Explanations other than the scleral-facial mismatch hypothesis have to be considered. For example, all non-human face stimuli with (originally) white sclerae are artificial reconstructions of humanity’s ancestors and relatives, whereas most faces of the type non-human with (originally) dark sclerae are photographs of ape faces. Therefore, the difference in likeability between those two types could be due to a dislike for (visibly) artificial faces, which would be expected based on existing uncanny valley effect research on artificially generated faces (see for example MacDorman & Chattopadhyay, 2016). An unexpected secondary finding was overall lower likeability ratings for human faces, compared to non-human faces. This is not in line with previous uncanny valley effect research, which consistently found the highest likeability ratings for the highest levels of human likeness. Based on this, congruent human faces (so, with white sclerae) would be expected to receive the highest likeability ratings, not congruent non-human faces (so, with dark sclerae) as we found. This suggests that there are different factors that influence likeability ratings, other than the combination of face morphology and scleral coloration. The heterogeneity of the stimulus set might have played a role in this disparity of ratings. While the (original) set of congruent non-human faces (so, faces with dark sclerae) contained, as mentioned above, mostly photographs of apes (orangutans, gorillas, chimpanzees) and contained only two reconstructions (of Australopithecus), the set of congruent human faces contained photographs of humans (Homo Sapiens), as well as 11 reconstructions (of Neanderthals and Homo Erectus). As mentioned above, the general dislike for artificial human faces might have led to lower likeability ratings for these reconstructions, independent of their face morphology and scleral coloration, thereby lowering the ratings of the overall face type. Yet another alternative explanation could lie in the classification of the stimuli. Classifying each stimulus as having a human- or non-human face morphology was done by the researcher. However, deciding whether a face appears overall "human-like" is not universally agreed upon, as can be seen in studies that employ a forced-choice categorization task (see for example da Fonseca Grebot et al., 2022). So, participants might have "disagreed" with the researcher-made classifications and might still have experienced these faces as mismatched. For yet another alternative explanation, there are differences between homo sapiens eye regions and eye regions of the other listed hominids, for example the pronounced brow ridge which homo sapiens lacks. Features like these might have led to lower likeability ratings, independent from differences attributable to scleral coloration. The inconsistent impact of the manipulation on the likeability of non-human faces makes it difficult to rule out any alternative explanation.

In conclusion, the results we found for human faces support the scleral-facial mismatch hypothesis, the results we found for non-human faces do not.

Eye Tracking Results

The main goal of the eye tracking component of the present research was to determine whether there are differences in gaze behaviour between regarding faces that cause the uncanny valley effect and faces that do not. More specifically, we investigated any possible differences in fixation count, dwell count, and hit time, on the entire face, and the main inner facial features, the eye region, the nose, and the mouth, between faces with- and faces without a scleral- facial mismatch.

We found that, across all conditions, the eye region was the most fixated inner facial feature. This is in line with existing literature that place the eyes as the most important facial feature for any task involving looking at a human face (Henderson et al., 2005; Itier et al., 2007; Janik et al., 1978; Laughery et al., 1971; Luria & Strauss, 2013; Philippe G Schyns et al., 2002). M. Cheetham et al. (2013), whose stimuli were morphs between artificial- and real human faces, found increased visual attention being directed towards the eye and mouth area for “boundary faces”. In the context of their study, these were faces that participants had difficulty classifying as human or non-human, which are the most likely to fall into the uncanny valley. da Fonseca Grebot et al. (2022) were not able to replicate these findings and instead found the nose region to receive increased attention in boundary faces. They explain this difference through a difference in experimental setup: where M. Cheetham et al. (2013) presented a centralized fixation point before each (equally centred) stimulus, da Fonseca Grebot et al. (2022) used a fixation point that was horizontally lateralized and presented randomly on either side before each stimulus. They did so to avoid a bias towards any specific AOI, especially the central one, that of the nose. This however had the opposite effect from what they intended, since participant had to perform an eye movement every time before viewing the stimulus, and the nose might serve as a visual “anchor” for quickly moving faces, being fixated more (Võ et al., 2012). In the present study, we were able to partially replicate the findings of Cheetham et al. (2013). For human faces and non-human faces, respectively, mismatched faces received more fixations on the eye region than congruent faces. We did however not find that pattern for the mouth AOI, where instead, mismatched faces were fixated less than congruent faces. We also found more fixations and dwells on the mouth AOI for non-human faces, as well as lower hit times, suggesting that this area got more visual attention than on human faces. Unlike da Fonseca Grebot et al. (2022), we found no significant increase in fixation count towards the nose for mismatched faces (which are the present study’s counterpart to their “boundary” faces) on the population level. However, individual differences in hit time suggested that some individuals did fixate the nose AOI on mismatched human faces before any other facial feature, which is partially in line with the findings of da Fonseca Grebot et al. (2022). We found a large spread of individual differences in fixation count on the eye region and on the nose, across all face types and both conditions, which might further explain the apparent differences between our findings and the findings of existing research discussed above, as well as the differences among the findings of said existing research. We did not use any fixation points before stimulus presentation, but since the button to progress from the rating scale slide to the next stimulus presentation was at the bottom of the screen, participants that did not use the spacebar to move to the next slide were most likely looking towards the bottom of the screen at stimulus onset and had to re-direct their gaze towards the center of the screen, where stimuli were consistently presented. This might have introduced a center bias towards the nose, similar to case of da Fonseca Grebot et al. (2022), which could potentially explain why the mouth AOI got the least visual attention. However, the overall distribution of gaze (or the hierarchy as the authors call it) of eyes, then, significantly lower, the nose, then the mouth, was the same in the present study as in the existing literature, so if the present study did introduce a center bias here, it was not large enough to overpower this overarching gaze distribution pattern. We also, as described above, did not find an increase in fixations towards the nose for “boundary faces” (which in our case would mean faces with a scleral-facial mismatch), which according to da Fonseca Grebot et al. (2022) would be expected if this bias had been present. We also found that the order in which participants first looked at all the facial features, measured by hit time, was consistent throughout all conditions and followed the same pattern as fixation count, with the eye region being hit first, then the nose, then the mouth. This could be the case because the eye region AOI was often the biggest AOI of a stimulus. However, this hit sequence result is in line with existing literature from social cognition research (Itier & Batty, 2009; Schyns et al., 2007), but has not been replicated in the context of the uncanny valley before. While the sequence was consistent, we did find a significantly reduced hit time on the mouth AOI for non-human faces compared to human faces. In combination with the increased fixation count towards this region for non-human faces, it seems like these features attract more visual attention than the same of human faces. While fixation count for the entire face was similar between human- and non-human faces, the distribution of visual attention across facial features differed. The eye region of human faces received more attention, whereas the nose and mouth area of non-human faces received more visual attention.

This is in line with existing research, which shows that human eyes attract visual attention more so than non-human primate eyes (Dupierrix et al., 2014; Emery, 2000).

In summary, answering the central eye tracking question of the present study, we found that the eye region of mismatched faces received more visual attention. The large spread of individual differences in gaze behaviour, however, should caution against over-generalizing these findings.

Strengths and Limitations

Since Koopman (2019) found the uncanny valley effect to be universal, we assumed the sample size and characteristics to be appropriate for the present study. Because investigating the uncanny valley effect as it applies to biological faces is by itself a new line of research, the present study was the first one to do so using both photo manipulation, as well as eye tracking measures. Investigating the uncanny valley effect using eye tracking has also not yet been done extensively and the present study was therefore the first to include dwell count and AOI hit time in the analysis, this analysis also being the first in this particular line of research to make use of multi-level modelling for analysing eye tracking data. Even though the main hypothesis of the present study was not entirely supported by the findings, the findings do suggest that the evolutionary approach to explaining the uncanny valley effect seems to be the correct and promising line of research.

Stimulus manipulation, as discussed above, was not sufficiently effective to isolate the effect of scleral coloration on likeability scale response data. This limits the degree to which the results can support the scleral-facial mismatch hypothesis. Existing uncanny valley eye tracking research used pupil dilation variation as a proxy measure for the eeriness response (Saneyoshi et al., 2022; Sharma & Vemuri, 2022), which was not possible to replicate with the setup of the present study, i.e., TobiiPro Fusion and iMotions.

While the low presentation time of two seconds per stimulus keeps both the amount of eye tracking data manageable, as well as the time requirement per participant down, it also meant that overall fixation count and dwell count, the main two variables used for eye tracking analysis, were rather low per participant and per AOI. Any differences between conditions were therefore also quite small, or possibly masked altogether. However, finding statistically significant differences at this scale, despite the aforementioned large individual differences, might mean that the effects are indeed behaviourally significant as well.

Future Research

The goal of future research should be to rule out alternative explanations that the present study left open, by further investigating the eye region mismatch hypothesis. Replicating the present study with improved manipulation seems advisable. This could be achieved by limiting the set of stimuli to the most realistic face reconstructions and letting an image processing expert handle the manipulation, altering not just scleral coloration but the entire eye morphology, to make it more or less human-like, respectively. Secondly, we recommend isolating the eye regions of the original stimuli, scaling them up and subsequently presenting them to participants to check whether presenting that part of the face is sufficient to trigger the uncanny valley effect. This could also enable the use of more fine-grain AOI, differentiating between brow ridge, sclera, and other features, to ultimately determine which exact combination of features is responsible for the effect. Lastly, to test whether the eye region is indeed sufficiently useful for species recognition and whether the uncanny valley effect is a deciding factor for boundary faces, a classification task, i.e., having participants decide whether a given eye region is human or non-human, could be added to such an experimental setup. This would also eliminate any potential researcher bias from face type categorisations.

Conclusion

The central question of the present research was to check our scleral-facial mismatch hypothesis: does the mismatch between scleral coloration and the humanness of a given face cause it to fall into the uncanny valley? We found that human faces with a scleral-facial mismatch are perceived as less likeable than scleral-facially congruent human faces. Scleral-facially mismatched faces also received more visual attention towards the eye region, compared to congruent faces. Since the pattern of reduced likeability for mismatched- compared to congruent faces was not found in non-human faces, findings from the present study do not fully support the scleral-facial mismatch hypothesis. The discovered effects are therefore not sufficient to explain the evolutionary origin of the uncanny valley effect. This is because relieving the scleral-facial mismatch by manipulation did not increase the perceived likeability in reconstructed hominin faces. Nonetheless, our findings suggest that the eye region is indeed the responsible area, but not the sclera by itself. We therefore encourage future research to further investigate the role of the eye region in causing the uncanny valley effect.

References

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: a reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 997.

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.

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 [Original Research]. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00108

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 [Article]. *Frontiers in Psychology*, *4*(MAR), Article 108. https://doi.org/10.3389/fpsyg.2013.00108

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

Ekman, P., & Friesen, W. V. (1978). Facial action coding system. *Environmental Psychology & Nonverbal Behavior*.

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/

Green, R. D., MacDorman, K. F., Ho, C.-C., & Vasudevan, S. (2008). Sensitivity to the proportions of faces that vary in human likeness. *Computers in human behavior*, *24*(5), 2456-2474. https://doi.org/https://doi.org/10.1016/j.chb.2008.02.019

Haeske, A. B. (2016). *The uncanny valley : involvement of fast and slow evaluation systems* [Bachelor Thesis, University of Twente]. http://essay.utwente.nl/69091/

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & cognition*, *33*(1), 98-106.

Ho, C.-C., MacDorman, K. F., & Pramono, Z. D. (2008). Human emotion and the uncanny valley: a GLM, MDS, and Isomap analysis of robot video ratings. Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction, 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.

Janik, S. W., Wellens, A. R., Goldberg, M. L., & Dell'Osso, L. F. (1978). Eyes as the center of focus in the visual examination of human faces. *Perceptual and Motor Skills*, *47*(3), 857-858.

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

Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2004). When Facial Attractiveness is Only Skin Deep. *Perception*, *33*(5), 569-576. https://doi.org/10.1068/p3463

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/

Langlois, J. H., & Roggman, L. A. (1990). Attractive Faces Are Only Average. *Psychological Science*, *1*(2), 115-121. https://doi.org/10.1111/j.1467-9280.1990.tb00079.x

Laughery, K. R., Alexander, J. F., & Lane, A. B. (1971). Recognition of human faces: effects of target exposure time, target position, pose position, and type of photograph. *Journal of Applied Psychology*, *55*(5), 477.

Lewis, M. B., & Edmonds, A. J. (2003). Face Detection: Mapping Human Performance. *Perception*, *32*(8), 903-920. https://doi.org/10.1068/p5007

Luria, S. M., & Strauss, M. S. (2013). Comparison of eye movements over faces in photographic positives and negatives. *Perception*, *42*(11), 1134-1143.

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

MacDorman, K. F., Green, R. D., Ho, C.-C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in human behavior*, *25*(3), 695-710. https://doi.org/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/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/https://doi.org/10.1016/S1364-6613(02)01903-4

McKelvie, S. J. (1976). The Role of Eyes and Mouth in the Memory of a Face. *The American Journal of Psychology*, *89*(2), 311-323. https://doi.org/10.2307/1421414

Mori, M. (1970). The uncanny valley: the original essay by Masahiro Mori. *IEEE Spectrum*.

Rhodes, G., Proffitt, F., Grady, J. M., & Sumich, A. (1998). Facial symmetry and the perception of beauty. *Psychonomic Bulletin & Review*, *5*(4), 659-669. https://doi.org/10.3758/BF03208842

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.

Schyns, P. G., Bonnar, L., & Gosselin, F. d. r. (2002). Show Me the Features! Understanding Recognition from the Use of Visual Information. *Psychological Science*, *13*(5), 402-409.

Schyns, P. G., Petro, L. S., & Smith, M. L. (2007). Dynamics of visual information integration in the brain for categorizing facial expressions. *Curr Biol*, *17*(18), 1580-1585. https://doi.org/10.1016/j.cub.2007.08.048

Seyama, J., & Nagayama, R. S. (2007). The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces. *Presence*, *16*(4), 337-351. https://doi.org/10.1162/pres.16.4.337

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

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

Walker-Smith, G. J., Gale, A. G., & Findlay, J. M. (1977). Eye movement strategies involved in face perception. *Perception*, *6*(3), 313-326.

Yarbus, A. L., & Yarbus, A. L. (1967). Eye movements during perception of complex objects. *Eye movements and vision*, 171-211.

Zhang, J., Li, S., Zhang, J.-Y., Du, F., Qi, Y., & Liu, X. (2020). A literature review of the research on the uncanny valley. International Conference on Human-Computer Interaction,

Appendix A

Informed Consent Form

Please tick the appropriate boxes	Yes	No
Taking part in the study		
I have read and understood the study information, 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 only involves the recording of anonymized eye tracking data and questionnaire answers, which cannot be traced back to me in any way.	○	○
Risks associated with participating in the study		
I understand that taking part in the study involves the risk of being mildly unsettled by images.	○	○
Use of the information in the study		
I understand that information I provide will be used for the master thesis of L. Limmer, which could possibly be scientifically published afterwards.	○	○
I understand that personal information that can identify me, such as my name, will not be recorded or shared beyond the study team.	○	○
I agree to be audio/video recorded for the purposes of eye tracking.	○	○
Future use and reuse of the information by others		
I give permission for the data that I provide to be archived so it can be used for future research and learning.	○	○
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.	○	○

Signatures

________________________	________________________	__________
Name of participant	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

Contact details for further information: Leonard Limmer, l.limmer@student.utwente.nl