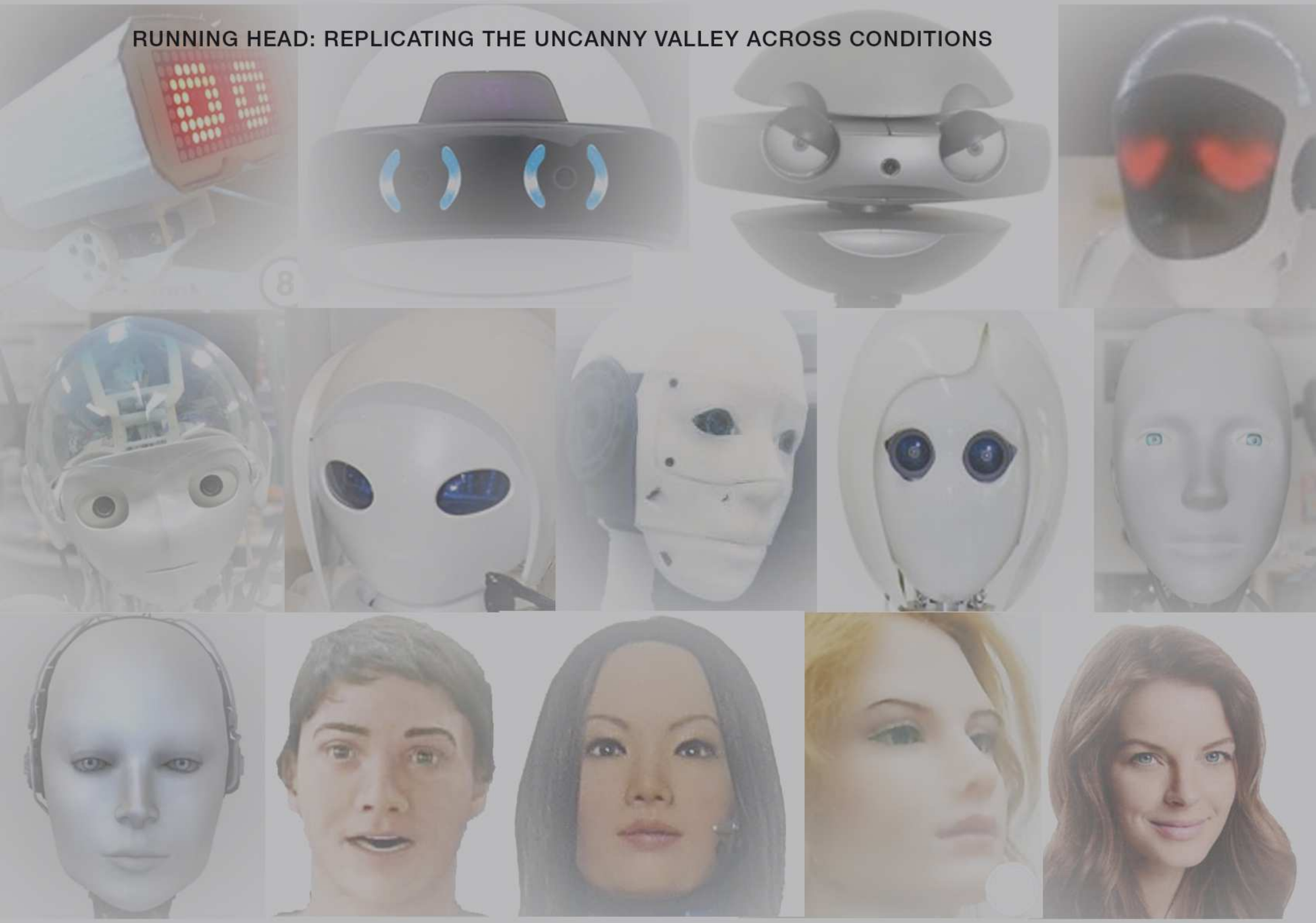

Replicating the uncanny valley across conditions using morphed and robotic faces

Master thesis Daan Keeris (s1024795)
Enschede, the Netherlands
November 2016

First supervisor: Dr. M. Schmettow
Second supervisor: Dr. M. L. Noordzij

UNIVERSITY OF TWENTE.

Abstract

The uncanny valley represents a strong dip in affect when observing stimuli with a high degree of human-likeness but that are not quite the real thing. This valley is thought to be caused by a mismatch between the human qualities that people are led to expect and the nonhuman qualities that are observed. Keeping in mind the recent replicability crisis, we attempted to replicate two contemporary studies on the uncanny valley. While doing so we expanded upon the studies in question in various ways in order to provide a better basis for future research. We incorporated stimuli with varying degrees of human-likeness. Half our stimuli were taken from another study and the other half were morphed faces of robots and humans. Using a within-subject repeated measures design, each participant rated the stimuli twice: once with a presentation time of 50 ms and once with 5 sec. Ratings were given based on an eeriness index. The results showed a clear replication of the uncanny valley phenomenon in the long condition. We were unable to reproduce a clear valley in the short condition however; instead we noticed a shift to the left on the human-likeness axis. Furthermore, we could not approximate the full curve with morphed faces as our stimuli only captured the right upward slope out of the valley. Based on these observations, which are partly in line with what we hypothesised, we attempt a cognitive theory on category confusion complemented by fluency of processing as an explanation of the emotional response and the valley's absence in the short condition.

Samenvatting

De ‘uncanny valley’ is een visuele representatie van een dip in affect wanneer men naar stimuli kijkt met een hoge mate van mensgelijkheid. Er wordt aangenomen dat deze dip veroorzaakt wordt door een discrepantie tussen de menselijke eigenschappen die men verwacht en de niet-menselijke eigenschappen die worden geobserveerd. Met het oog op de recente reproduceerbaarheidscrisis pogen we om twee recente onderzoeken over de uncanny valley te reproduceren. Tijdens dit proces hebben we de studies uitgebreid om zo een betere basis voor toekomstig onderzoek te kunnen vormen; we hebben stimuli met verscheidene gradaties mensgelijkheid gebruikt. Een helft kwam van een andere studie en de andere helft bestond uit gezichten die deels menselijk en deels robotisch waren. De participanten beoordeelden elke stimulus twee keer: eens met een presentatietijd van 50 ms en eens met een presentatietijd van 5 sec. De beoordelingen werden gegeven gebaseerd op een index speciaal gemaakt om ‘eeriness’ te meten. De resultaten lieten de uncanny valley duidelijk zien in de 5 sec conditie, maar niet in die van 50 ms. In deze conditie zagen we echter wel een verschuiving naar de linkerkant van de as die mensgelijkheid weergeeft. Verder waren we niet in staat om de volledige valley te schatten met de gemengde gezichten omdat onze stimuli enkel de rechterhelft van de valley bleken te dekken: de klim omhoog. Op basis van deze observaties, die ten dele overeenkomen met onze hypotheses, passen we een cognitieve theorie over ‘category confusion’ toe, aangevuld met ‘fluency of processing’. Op die manier proberen we de emotionele respons te verklaren alsook het gebrek aan de valley in the 50 ms conditie.

Contents

Introduction

Early in the 1990s, computer-generated imaging (CGI) started appearing more and more in movies; famous titles such as *Terminator 2* (1991) and *Jurassic Park* (1992) made use of CGI and did so with great success. The success of the first animated 3D movie released in 1995, Pixar's *Toy Story*, marked the beginning of mainstream animation production as technology became advanced enough to create realistic human characters. In 2001, these were shown in hyper-realistic fashion in the movie *Final Fantasy*. For example, one of the characters in the movie was animated in so much detail that it took an hour and a half for every frame in which she occurred (Brook, 2007). This time the expected success stayed off however. Movie critic Peter Travers stated that "At first it's fun to watch the characters ... [b]ut then you notice a coldness in the eyes, a mechanical quality in the movements" (Travers, 2001). It can be argued that the producers of *Final Fantasy* fell into the trap of "the uncanny valley". The aim of this study is to replicate two contemporary studies on this phenomenon in order to get a better view of when and under what conditions the uncanny valley takes place. This could be a take-home message for those involved in the design of artificial faces, so that they can better attempt to avoid any eeriness caused by their creations.

The Japanese roboticist Masahiro Mori (1970) first introduced the concept of the uncanny valley in a hypothetical graph plotting familiarity on the one axis versus human-likeness on the other (see Figure 1), with the goal of explaining why some faces may appear eerie to viewers. The non-linear graph shows that robots and other artificial characters or objects are rated more positively the more human-like their appearances become, up until the point where the robots are sufficiently realistic that the remaining non-human features become noticeable and disturbing (MacDorman & Entezari, 2015). This is where an observer may have difficulty distinguishing the object from its natural human counterpart (Cheetham, Suter, & Jäncke, 2011).

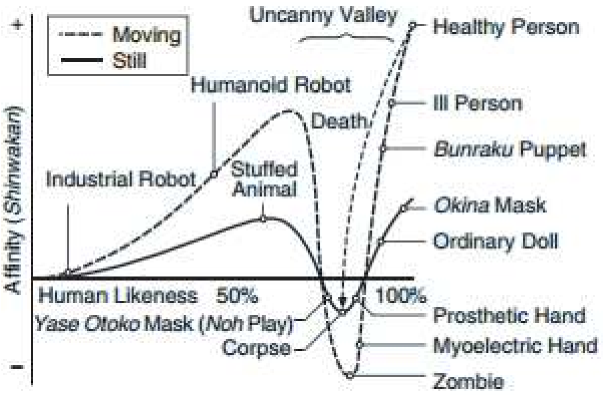

Figure 1. The non-linear graph depicting the uncanny valley (Mori, 1970)

The uncanny valley phenomenon refers to the feelings of eeriness and discomfort when confronted with these robots and other realistic human figures, with this dip in evaluation resembling the characteristics of a valley: a deep descend before climbing back up once appearances start to look like those of real human beings again. MacDorman, Green, Ho, and Koch (2009) state that this is because the nonhuman imperfections expose a mismatch between the human qualities that people are led to expect and the nonhuman qualities that are observed.

One goal of our study is to replicate an experiment by Mathur & Reichling (2016). They chose to present a sample of real-world android robots without any faces of real human beings involved, with the goal of determining if human reactions to faces of android robots indeed exhibit the uncanny valley phenomenon. A second goal is to replicate another study on the valley; that study instead used morphed faces of humans and robots rather than real-world stimuli.

Influencing factors on the uncanny valley

There are many factors that influence people's perception of artificial characters in such a way that they experience feelings of uncanniness. In his exposition of the uncanny valley Mori (1970) states that these feelings are dependent on the degree of familiarity, or to what extent the stimulus is considered human-like.

Facial features are a common influencing factor on the uncanny valley. For example, Seyama and Nagayama (2007) showed in their study that the phenomenon was especially noticeable when face images involved particularly bizarre features such as abnormal eye or head size. MacDorman et al. (2009) reported similar results in their experiments, with unrealistic facial proportions of still computer-generated faces leading to increased eeriness ratings. Another contributing physical aspect is the expression of certain emotions (i.e. fear, sadness, disgust, and surprise) in the upper part of computer-generated faces during speech, as these have an influence on the character's uncanniness as well (Tinwell, Grimshaw, Nabi, & Williams, 2011). Furthermore, the way people perceive experience is an important part of the uncanny valley according to Gray and Wegner (2012), who showed that perceptions of mind are linked to feelings of unease.

Hypotheses on the uncanny valley

MacDorman et al. (2009) divide possible explanations of the uncanny valley into two groups of hypotheses: those that involve a broad and general range of cognitive processing that occurs relatively late in perception, and those that involve automatic, stimulus-driven, specialised processing that occurs early in perception. Should the uncanny valley be explained using the first category, it is plausible that eeriness ratings are increased the longer an artificial face is observed due to the later involvement of higher cognitive processes. Moll (2015) has shown, however, that this is not the case; while 500 ms was a stronger predictor for an unlimited presentation time, at least part of the judgment took place in 50 ms. This implies that the uncanny valley phenomenon is likely to be explained using the second classification provided by MacDorman et al. (2009). For this category, where automatic and specialised processing plays a key role, they provide three specific explanations with the first one being threat avoidance. This hypothesis is based on the theory of disgust by Rozin and Fallon (1987) which considers disgust to be an evolved cognitive mechanism to make sure that humans avoid infection. To illustrate: while a healthy human being generally does not elicit disgust, a diseased person (e.g. a leper) or one with visible gene defects is more likely to be perceived as eerie. MacDorman and Ishiguro (2006) reason that, although eeriness and disgust are not the same, the situation may still be applicable to the uncanny valley by having eeriness perform the same function disgust does in Rozin and Fallon's theory. Furthermore, Öhman (2000) found that perceived defects in characters with a large degree of human-likeness can elicit aversion motivated by fear and consequently trigger a fight-or-flight response. The second hypothesis discussed by MacDorman et al. (2009) is regarding shared circuits for empathy. It is based on research indicating that perceptual, cognitive, and affective processing may work in tandem when perceiving uncanniness (Chaminade, Hodgins, & Kawato, 2007; Krach et al., 2008). These processes are what is referred to as 'shared circuits' and support humans in their ability to understand intentions, as they are active both when a person performs an intentional action as well as when a person sees someone performing that same action (Keysers & Gazzola, 2007). However, Krach et al. (2008) used functional magnetic resonance imaging to show that the shared circuits are not only activated when observing other humans, but increase linearly based on human-likeness. This entails that watching robots activates the same regions. With humans and robots activating the shared circuits of empathy when performing similar actions, it is plausible that feelings of uncanniness emerge when observing human-like robots. The third and last hypothesis that can be considered to fall under the second category of hypotheses according to MacDorman et al. (2009) is in regards to evolutionary aesthetics. Many studies correspond to this categorisation, with results showing that the perception of attractiveness has a biological basis in automatic, stimulus-driven, specialised perceptual processing (for a meta-analytical review see Langlois et al., 2000). This argumentation is further supported by research indicating humans' high agreement on the assessment of attractiveness (Olson & Marshuetz, 2005; Willis & Todorov, 2006). In assessing attractiveness, features that were perceived as particularly attractive were various signs of reproductive fitness (e.g. Soler et al., 2003). By extension, perceiving those who could be considered unattractive as lacking in reproductive fitness may have led to the evolution of the same cognitive mechanisms that are involved in experiencing feelings of eeriness in the uncanny valley.

To summarise, there are two categories of hypotheses that could explain the uncanny valley phenomenon. The first one relies on a broad and general way of cognitive processing and involves more conscious reflection of what is observed. This type of reflection, however, should require longer processing times when forming a judgment of eeriness. The second group of hypotheses focuses on forming judgments fast and automatically without needing to consciously reflect upon what is observed. Research has shown that processing faces falls within the latter category and as such is a specialised, fast, and automatic process.

Face processing

Because face processing plays an integral role in our study, some information on processing and presentation times is important. Several studies using presentation times ranging from 17 ms to 50 ms have indicated that this range is enough to process faces depending on the context and task. For example, in a study by Stone, Valentine, and Davis (2001) participants were able to subconsciously discriminate the affective valence of famous or familiar faces with better than chance accuracy; they could categorise faces into good or evil after a processing time of a mere 17 ms, with a pilot study showing higher accuracies for 33 ms and 50 ms. Stone et al. (2001) opted for 17 ms to ensure that it would be unambiguous of prime faces used in the study, but also to ensure that the prime faces would still be perceptible. A large number of the faces used in the experiments of their study were rated as very familiar, however, which may have skewed the data. Mogg and Bradley (1999) found evidence that facial expressions can be unconsciously recognised at a presentation time of 14 ms, which also suggests that faces are perceptible at a duration of 17 ms. Another study where participants needed a minimum presentation time of 17 ms was done by Grill-Spector and Kanwisher (2005); their research showed that participants were able to categorise objects with a presentation time of 17 ms as faces, although accuracy was low. When the exposure time was increased to 33 ms and 50 ms however, accuracy increased to nearly 50% and more than 70%, respectively. These results combined with the promising pilot study by Stone et al. (2001) suggest that a presentation time of 33 ms is suitable for future research. Furthermore, another study has demonstrated that observers are able to form consistent first impressions on threat judgments based on information presented in 39 ms presentations, whereas a presentation time of 26 ms was too little to do so (Bar, Neta, & Linz, 2006). In another study, Moll (2015) measured the actual time participants needed to form a stable judgment of uncanniness through the processing of human faces. Here it was shown that the time needed to form judgments of uncanniness was comparable to the times found to judge human faces regarding the level of threat experienced; participants needed more than 17 ms while 50 ms was enough. This implies that the exact time needed to decide whether a face is perceived as uncanny is somewhere between 17 and 50 ms (Moll, 2015). This corresponds with the results of other studies that used similar presentation times.

Summarised, the short literature review above on presentation times of both human and artificial faces shows that face perception is indeed an automatic and specialised process, therefore categorising it under the second group of hypotheses as discussed by MacDorman et al. (2009). These specialised processes allow humans to form judgments automatically based only on the visual appearance of a face, with the level of uncanniness being one of the judgments in question. This implies that it is possible to judge how long it takes for humans to form a judgment on uncanniness, and research by Moll (2015) has shown that this time lies within the 17-50 ms range. Therefore a second goal of this study is to partly replicate Moll's (2015) study. This experiment should show whether human reactions to android robots indeed exhibit the uncanny valley phenomenon when using robotic faces. Additionally it should show at what point the increase in perceived eeriness takes place.

Individual differences

In their study, MacDorman and Entezari (2015) attempted to explore how differences among individuals could affect sensitivity to the uncanny valley. They did this by determining the relation between personality traits and their emotional state while perceiving potentially uncanny androids, with androids being defined as “very humanlike robots” (MacDorman &

Ishiguro, 2006, p. 298) in order to distinguish them from the more mechanical-looking humanoid robots. MacDorman and Entezari (2015) chose to operationalise uncanny valley sensitivity as higher ratings of eeriness and lower ratings of warmth for androids, in accordance with the scale developed by Ho and MacDorman (2010). The following nine trait indices were tested: perfectionism, neuroticism and anxiety, personal distress, animal reminder sensitivity, human-robot and android-robot uniqueness, religious fundamentalism, and negative attitude towards robots. These traits were motivated by proposed theories of the uncanny valley (MacDorman & Entezari, 2015). Of the nine indices tested, only perfectionism and personal distress had no significant correlation. That is, they were associated with android eerie ratings but not with warmth (r=0.01 and r=-0.07, respectively). Furthermore, eight of the nine traits had significant positive correlation with android eerie ratings, with android-robot uniqueness being the exception (r=-0.03). Additionally, the study showed that both age and gender correlated significantly with android eerie ratings as well, with females and younger undergraduates rating the androids eerier. Ho et al. (2008) came to the same conclusion; in their study, younger female participants considered robots to be eerier than did the male participants.

Societal impact and application domain

It is important to evaluate the importance of the uncanny valley in contemporary and future society for several reasons. First, it is likely that the future brings increased exposure to artificial characters; robots, for example, are already involved in elderly care (Wada, Shibata, Asada, & Musha, 2007) and child’s play (Iromec, 2009), and it is likely that their use in these and other contexts will become more ubiquitous in the future (Bemelmans, Gelderblom, Jonker, & de Witte, 2012; Ho, MacDorman, & Pramono, 2008; Tapus, Ţăpuş, & Matarić, 2008). Even more so because each new generation of robots comes progressively closer to simulating real human beings in appearance, facial expressions, and gestures (MacDorman et al., 2005; Matsui, Minato, MacDorman, & Ishiguro, 2005; Minato, Shimada, Ishiguro, & Itakura, 2004). There is also the practical benefit that a better understanding of the uncanny valley may lead to the development of more effective artificial characters. After all, these are often being used in perceptual and social interaction studies (e.g. Bailenson & Yee, 2005; Boker et al., 2011; Von der Pütten et al., 2010). Additionally, a survey by the Entertainment Software Association (2015) showed that 155 million people play video games in just the US; a number that is not only indicative of large exposure to artificial characters, but also of the use of computer-generated animations. This leads to a second reason why the uncanny valley is important: several factors influence concerns about the uncanny valley in regards to the ever increasing use of computer-generated animation. These concerns have significant impact on both the animation as well as the video game industries (MacMillan, 2007). For example, a study has shown that hyper-real characters fail to establish an emotional bond with the audience due to the suppression of empathy (e.g. Butler & Joschko, 2009; Kaba, 2013). Some viewers fail to identify with the characters, which are perceived as soulless and vacant (MacDorman & Chattopadhyay, 2016). Scholars and film critics have also related the uncanny valley to viewers' dyspathy for 3D computer-animated characters that have a high degree of human-likeness (Butler & Joschko, 2009; Freedman, 2012). The result of this is that the title may flop with a possible outcome being the production studio going bankrupt (MacDorman & Entezari, 2015). Research has indicated that viewers have indeed reported uncanny experiences in response to realistic virtual characters (Tinwell, Nabi, & Charlton, 2013). To avoid such experiences, studios like Pixar opt for a more cartoony style of animation rather than human photorealism (Canemaker, 2004). This is in line with what children (both typically developing ones as well as children with autism) prefer: simplicity, exaggerated facial features and cartoon-like features trigger positive affect (Peca, Simut, Pintea, Costescu, & Vanderborght, 2014). To illustrate, CGI artist Jonathan Joly (2008) created an adaptation of the uncanny valley (see Figure 2). Here it is clear that most commercially successful animations all hit the high graph while the ones that flopped are all associated with feelings of unease.

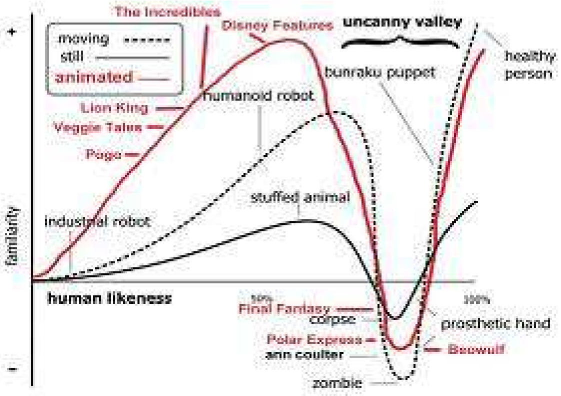

Figure 2. Adaptation of the original graph by Mori (1970) showing the position of several animations (Joly, 2008).

Objectives of this study

In light of the recent replicability crisis (Open Science Collaboration, 2015), the main objective of the study is replicating contemporary research. In order to do so, we will set up an experiment with several purposes. The first of these purposes, or goals, is to conduct a similar experiment as the one done by Mathur and Reichling (2016). It is our hypothesis that we will not only corroborate the finding that a judgment of eeriness can be seen during long presentation times, but during short presentation times as well. This because the uncanny valley is most likely to be explained by fast and automatic processes and because research on face processing and perception has shown that a very short presentation time is enough to categorise faces. Our experiment adds to that of Mathur and Reichling (2016) in two different ways: first, we make use of the eeriness scale by Ho and MacDorman (2010) whereas they used a different scale that was not specifically aimed at measuring eeriness. Secondly, our experiment has different conditions due to our manipulation of the presentation time. Whereas Mathur and Reichling (2016) used only a long condition, we use both a long and a short one. The short condition has a presentation time of 50 ms, based on aforementioned studies on face processing. Because we opt to use another scale than Mathur and Reichling (2016) to measure eeriness, a secondary goal is to compare both scales. This to determine which one is more effective in terms of measuring eeriness in the uncanny valley. The third goal is again a replication, this time of the study by Moll (2015). In our approach we partly replicate the studies that make use of morphed blends of human and robot faces, specifically the study by Moll (2015). The artificial faces used in that particular fashion in this study will have varying degrees of human-likeness based on the morphing levels. The results of this study can serve to expand the basis for future research on the uncanny valley phenomenon, as well as assist in finding the most optimal artificial faces for more practical applications such as movies, games, toys, and elderly care.

Method

Participants

The sample used in this study consisted of 35 participants (29% female) with a mean age of 26. The majority of participants were native Dutch speakers (77%) with the remaining ones being native German speakers. More than half of the participants (69%) were recruited using the researcher's social network and did not receive any form of compensation. The remaining ones were found through the University's recruitment system; these participants were undergraduate students and received credits for participating in our study. All participants declared voluntary participation by signing an informed consent. Furthermore, this study was approved by the ethical committee of the Behavioural, Management, and Social Sciences faculty of the University of Twente.

Materials

The experiment took place in a 4m^2 laboratory provided by the University of Twente. The room contained a desk and chair, which was placed at a distance of approximately 75 cm from the screen. The computer used in our study had an Intel Core i7-3770 CPU, 8 GB of RAM and used Windows 7 64bit as operating system. The monitor on which stimuli were presented was a 22" LG E2210 with a refresh rate of 60Hz and 5 ms response time. A standard mouse and keyboard were supplied. For this experiment we used two different programs. PsychoPy (v.1.83) was used in order to run the experiment itself, and we used FantaMorph to merge the robotic faces supplied by Mathur and Reichling (2016) with the human faces used in this study. The human faces were retrieved from the database of the European Conference on Visual Perception (2008). By merging these faces in FantaMorph we were able to adjust the degree of human-likeness. The stimuli provided by Mathur and Reichling (2016) that we used included numbered circles. We felt that these could prove to be a distraction for participants, so Photoshop was used to edit them out. Photoshop was also used to slightly modify some of the robot faces that were merged with those of a human in order to make them more suitable for the morphing process. To measure judgments of eeriness we used the questionnaire developed by Ho and MacDorman (2010) translated to Dutch and German.

Study design

We used a within-subjects repeated measures design where each participant rated the same stimulus on two occasions using different presentation times. The independent variable here is the presentation time of the stimuli used in the experiment. We used two variations: 50 ms and 5 sec. The dependent variable was the eeriness rating as judged by the participants after the stimulus had been presented to them.

Stimuli

In order to replicate contemporary research we incorporated stimuli with varying degrees of human-likeness. This variety was achieved using existing stimuli from another study as well as using morphing software to merge a picture of a human face with a picture of a robotic face. Using the software we could freely adjust the ratio to which each face was represented, with the total always being at 100% (e.g. a resulting morphed face could look 21% human and 79% robotic). The robotic faces were retrieved from the study by Mathur and Reichling (2016). They used 80 faces in their experiments and we picked half of them using a systematic sampling method; every second robotic face was selected as a stimulus for our experiment.

The remaining 40 stimuli used in our study were created using four faces of real human beings and four robotic faces retrieved from a systematic image search on Google (for in- and exclusion criteria see Mathur and Reichling, 2016). Each human face was then morphed with a robotic face using FantaMorph a total of eight times. This means that one sequence consisted of ten different faces: both original faces plus the eight morphs. Each morph varied in their human-likeness with irregular increments; the changes in morph percentage were smaller there where we expected to especially notice the uncanny valley effect (see Appendix A for an overview of the morphing percentages per sequence). For an example of such a sequence, see Figure 3. With each sequence consisting of ten stimuli and there being four sequences in total, we created 40 different stimuli. Combined with the 40 robotic faces retrieved from Mathur and Reichling (2016), the total amount of stimuli used in this experiment was set at 80.

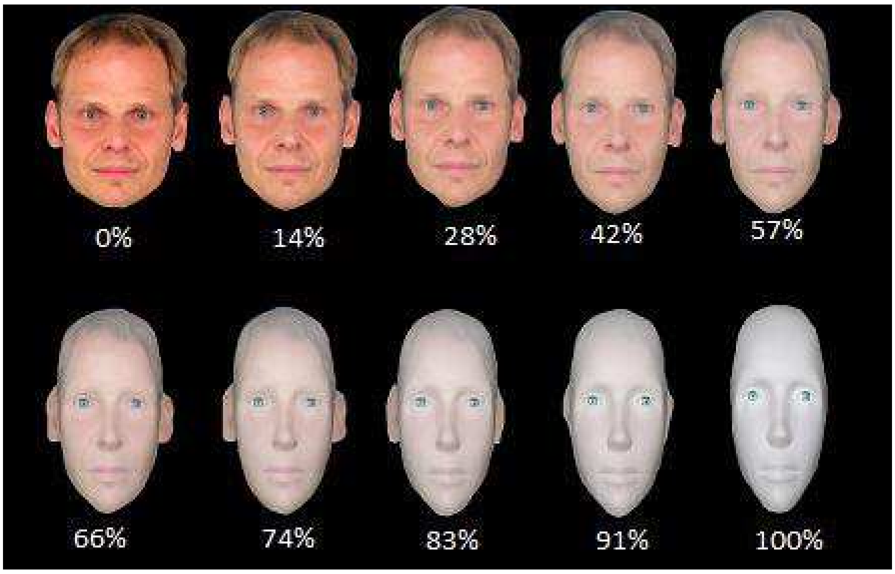

Figure 3. Example of one of the morphing sequences used in this study, showing the transition from a genuine human face (0%) to a fully robotic one

Ratings

The scale participants used to rate the stimuli was a questionnaire by Ho and MacDorman (2010), specifically the scale that measures the eeriness construct. This scale consists of eight different items measuring to what extent participants consider a stimulus as eerie. This scale was preferred for our research because alternative measurement instrument, such as the 'Godspeed Index' (Bartneck, Kulić, Croft, & Zoghbi, 2009), that were used in similar experiments were not suitable. The Godspeed Index used five different indices (anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety) to evaluate reactions to artificial characters. However, Ho and MacDorman (2010) judged this scale as unsuitable for investigating the uncanny valley due to none of these indices specifically corresponding to eeriness, a dimension that cannot be ignored when evaluating whether an artificial stimulus lies within the uncanny valley or not. They also criticised the indices themselves, stating that in the development of these indices there had been no attempts of making them de-correlate from positive (vs. negative) affect or even from each other. The eeriness index of the scale used in this study, however, is de-correlated from the humanness, warmth, and attractiveness indices developed by Ho and MacDorman (2010).

The original questionnaire by Ho and MacDorman (2010) is in English. However, we provided translations of the scale in both Dutch and German to pre-emptively prevent any misunderstandings due to a potential language barrier. The items of the original scale were translated from English to Dutch and German by a native speaker of the respective language, after which this translation was then given to another native speaker who in turn translated it back to English (see Appendix B for an overview of all original items and their translations). This way we minimised the possibility of faulty or inconsistent translations.

The item-stimulus pairing was randomly selected for each participant, resulting in ratings for all eight items on the scale.

Item-stimulus pairing

The program required to run the experiment randomly paired each stimulus to one of the eight possible items from the scale. The coupling remained during the experiment, meaning each stimulus was paired with the same item in both conditions. Consequently, each participant rated a specific stimulus on the same item for both of the presentation times. While the item-stimulus pairing was set as soon as the experiment began, the order in which the stimuli were presented was not. This means that a stimulus presented early in the 50 ms condition could be presented in the middle or at the end of the 5 sec condition. This lack of ordering minimised any order effects that could potentially affect the study results.

Procedure

Prior to the start of the experiment participants were given an informed consent form. Upon signing and therefore agreeing to the informed consent, they sat down in the experimental room and were given the opportunity to learn more about the motivation behind and the goal of the study. They were allowed to ask questions before starting the experiment and reassured that they were allowed to withdraw for any reason. Upon starting the experiment, the participants were shown several practice trials using the faces of famous individuals. These trials allowed the participant to practice and get a general feel of how the experiment would proceed. During the practice trials a researcher was present to answer any questions the participants might have regarding the procedure. Once the practice had been completed, the researcher left the room and participants could begin the actual experiment.

Afterwards, they were given the opportunity to ask further questions and to sign up for a mailing list in order to be updated with the results of the experiment once data analysis had been completed.

Task

Once the trials started, the participants were first presented a black screen for 500 ms. This screen was followed by a fixation cross, also for 500 ms. After the cross, the stimulus itself was presented for either 50 ms or 5 sec, depending on what phase of the experiment the participant was in at that time. Following the stimulus was a mask with the purpose of inducing a conflict in the stimulus perception; the processing of the stimulus is interrupted by the mask before the stimulus can be fully processed. This conflict does not affect the early stages of processing; it merely involves a competition for the higher level mechanics that are involved in object recognition (Kolers, 1968). This reduces the amount of higher level processing that takes place after stimulus presentation and consequently this should result in responses that are influenced by the processes that take place during the actual stimulus presentation. After the mask had disappeared a rating scale appeared where participants could rate the eeriness of the stimulus using a visual analog scale. This type of scale can provide more precise and psychometrically valid ratings than a Likert-type ordinal scale (Reips & Funke, 2008). The rating process was done by using the mouse to move the cursor anywhere (i.e. the participants were not restricted to whole numbers) on the scale and clicking the left mouse button. The full process of stimulus presentation can be seen in Figure 4 in the form of a flow chart.

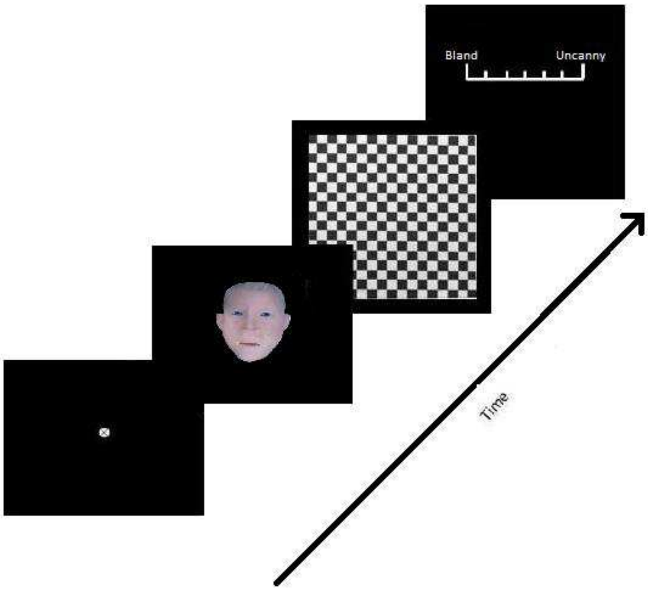

Figure 4. Flow chart of a stimulus as presented in the experiment

Data analysis

It is important to add that the responses in both experiments are on different scales. On the one hand there is the one-item likeability scale used by Mathur and Reichling (2016), and on the other hand there is the eeriness scale by Ho and MacDorman (2010) that we opted to use. Therefore an added objective of the study is to compare both scales to see which one is more suitable and why. We have done some harmonisation by bringing both scales to the [-1;1] interval and reversing the direction of the eeriness items. Hence, high responses denote a positive emotional response to the stimulus. Additionally, the human-likeness score of Mathur and Reichling (2016) has been normalised to the [-1;1] interval, where high values correspond to a stronger human-likeness of the stimulus. Furthermore, the morphing levels of our own stimuli were originally indexed from 1 to 10, with low numbers denoting human-likeness instead. This has been harmonised with the huMech score used by Mathur and Reichling (2016) in direction only. That is, the full range was scaled to the [0;1] interval so that high numbers denote strong human-likeness as well. We did not scale it to [-1;1] because the robot target faces were selected from the middle range of Mathur and Reichling's (2016) set.

For our regression analysis, we start with the same model as Mathur and Reichling (2016), using a third degree polynomial on averaged data: $f(x)=b0+\beta1x+\beta2x^2+\beta3x^3$ (see Appendix C for the function we used to compute the trough of our third degree polynomials). We opt to use a third degree polynomial to stay as close as possible to the analysis done by Mathur and Reichling (2016). They performed several F-tests to compare models of the second, third, and fourth degrees and concluded that the third degree model fit significantly better than the other two. As we aim to replicate their results, we will use the above polynomial for our data analysis. This polynomial leads to two new variables, huMech2 ($huMech^2$) and huMech3 ($huMech^3$), alongside the initial huMech0 and huMech1. Starting with a third degree polynomial regression we only have fixed effects due to the averaging over participants.

Results

Using our third degree polynomial we estimated the averaged data over stimuli. Table 1 shows these estimates as well as the deepest point of the curve across all variables for each condition. The troughs were found by calculating at what points the derivative of the polynomial is equal to 0 so that we could find the local minima. In order to calculate these points we first had to derive the polynomial once (first derivative), and from this derivative we were able to find the two points where it hits 0. Since we then did not yet know whether the points are local extrema or not, we had to derive once more (second derivative). Plugging our two points into the second derivative, we got positive values which ensured that those points were local minima. Table 2 provides an overview of the confidence intervals for all three conditions used in our study.

Table 1. *Estimates of averaged data over participants and the trough across all variables for each condition using a third degree polynomial*

Condition	β0 (huMech0)	β1x (huMech1)	β2x^2 (huMech2)	β3x^3 (huMech3)	Trough
Keeris_long	-0.136	-0.162	0.310	0.290	0.203
Keeris_short	-0.232	0.242	0.237	0.012	-0.532
Mathur_long	-0.203	-0.538	0.294	0.786	0.369

Table 2. *Confidence intervals of the trough position for each of the three conditions*

Condition	Center	Lower bound	Upper bound
Keeris_long	0.244	-0.316	0.468
Keeris_short	-0.482	-1.123	0.080
Mathur_long	0.366	0.293	0.452

In Figure 5 we can see that the shape of the fitted curve seen in our long condition is similar to the results of Mathur and Reichling (2016). In both graphs we see a clear uncanny valley curvature with a noticeable trough, thereby confirming the existence of the actual phenomenon. Several other interesting observations can be made when looking at Figure 5.

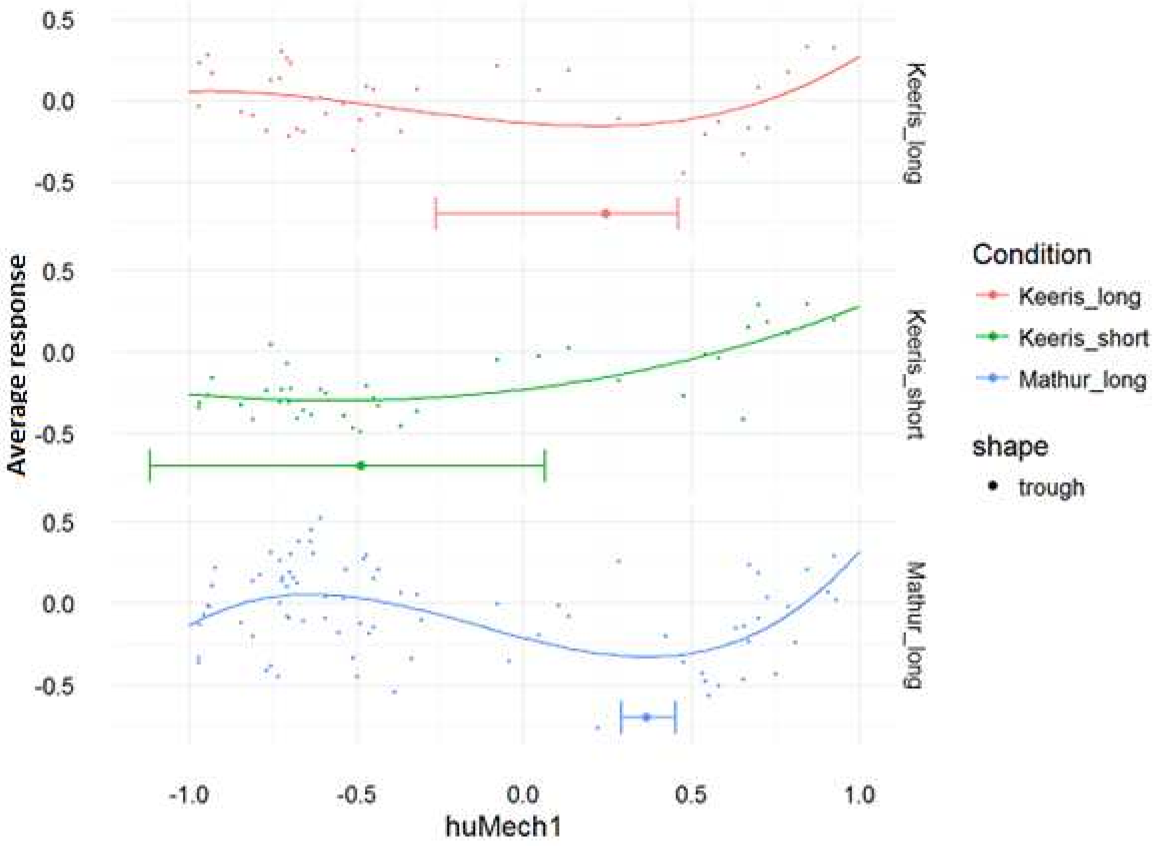

Figure 5. The fitted curves of all three conditions with error bars representing standard deviation of the mean

First, both long conditions have similar trough positions, though a visual check shows that ours is a little more orientated towards the mechanical pole of the scale. Note that we can only say that with a high degree of uncertainty, however, leading us to a second finding. The trough estimates of our data set have a much higher degree of uncertainty than that of Mathur and Reichling's set. This disparity seen in the lengths of the error bars can be attributed to our lack of data points around the estimated trough positions. Mathur and Reichling (2016) used many more participants in their study (n=342) where each participant received a selection of faces to judge. Each face was rated by 64 participants on average whereas in our study this was only 35. Thirdly, while we can clearly see the valley in both long conditions, the expected curve in the short condition is conspicuous by its absence. The lack of a clear uncanny valley curvature in the condition with a 50 ms presentation time means we cannot see nor replicate the uncanny valley in a condition with a short presentation time. In fact, it is not certain that it is a trough at all; due to its ambiguous shape it is also possible that it is, for example, a monotonously rising curve instead. The curves of both long conditions at least have a clear trough, despite them not being as deep as in the original depiction of the uncanny valley. In correspondence with Figure 5, Figure 6 shows the individual differences for both conditions. We see that for nearly every participant the short condition lacks a clearly defined trough where expected based on the original depiction of the valley. For the long condition however there are numerous participants (e.g. participants 7, 22, 33, 34) where the position of the curve and its trough correspond to Mori's (1970) or Mathur and Reichling's (2016) graphs.

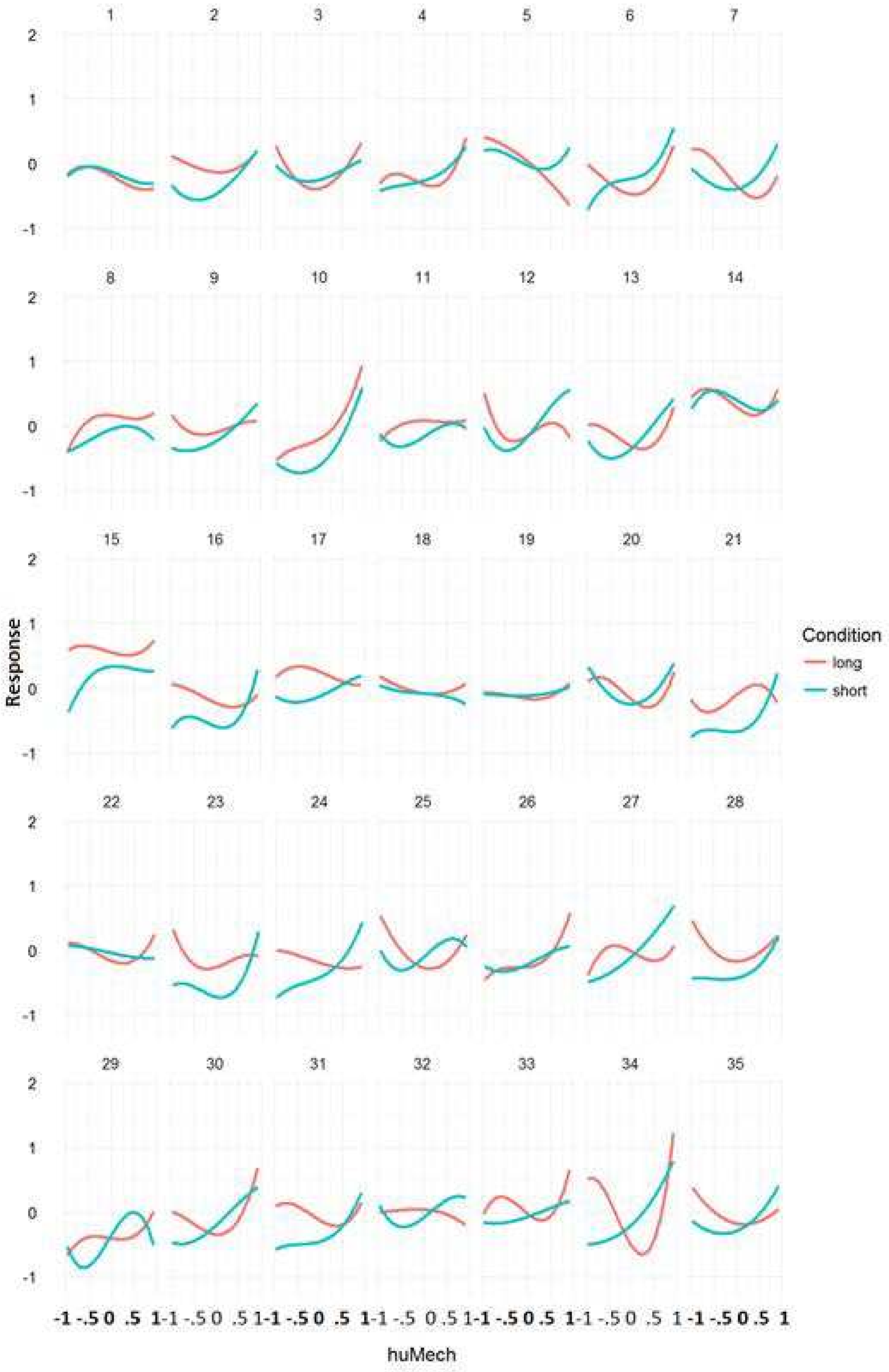

Figure 6. Individual differences of own data smoothened by a third degree polynomial

Results also indicate that we failed to approximate the full curve with morphed faces. Using our stimuli we only managed to capture the right upward slope out of the valley with the sequences (see Figure 7). This is especially clear for sequences B and D. As a result we cannot estimate the position of the trough with any reasonable accuracy. Possible explanations for this observation will be discussed in the next section.

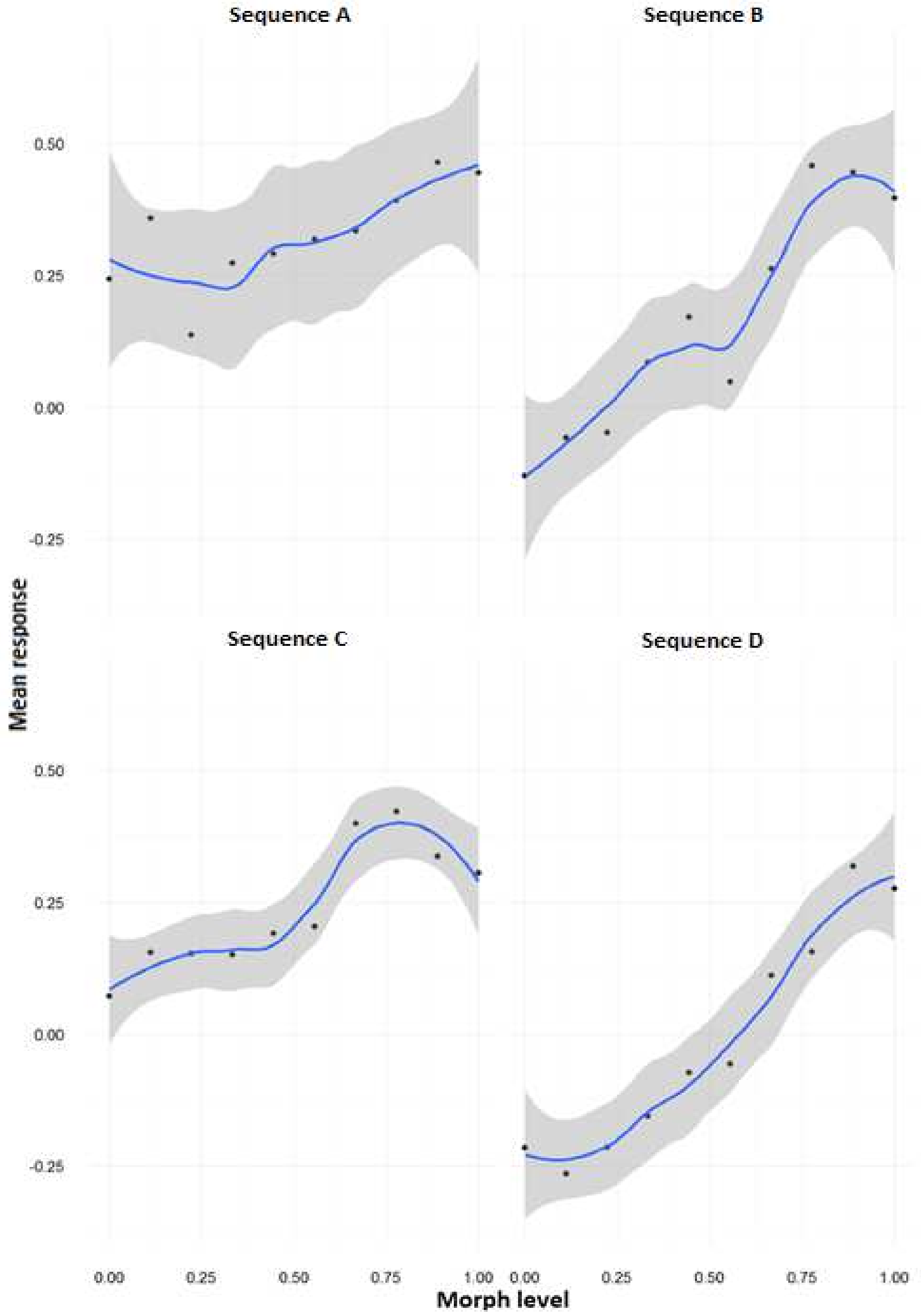

Figure 7. The mean responses at a long presentation time showing only the upwards slope for each sequence

Discussion

With this study we had several goals in mind. First, we aimed to replicate the study by Mathur and Reichling (2016) on the uncanny valley effect. They managed to capture the uncanny valley curvature by presenting mechanistic faces ranging from very robot-like, to faces with a high degree of human-likeness. In order to replicate their results we used half of their stimuli set. Our results show that while we were indeed able to see the uncanny valley effect in the long condition, we were unable to do so in the short condition. Consequently, we reject our initial hypothesis stating that we can see and replicate the uncanny valley phenomenon in a condition with short presentation times. A second goal of this study was determining the effectiveness of the eeriness index developed by Ho and MacDorman (2010). It turned out however that these scales were difficult to compare due to the large differences between the items because the items of one scale were much more extreme than items on the other scale. A third objective of our study was to partly replicate a study by Moll (2015). He used morphed artificial faces with varying degrees of human-likeness to determine that humans indeed experience the uncanny valley when looking at robots. However, we were unable to reproduce these results as well. That is, we could not approximate the full curve but only a part of it.

Research goals

For our first research aim we wanted to replicate the study by Mathur and Reichling (2016) while expanding upon it in two ways. One of these ways was the addition of a condition with short presentation times in order to give a better idea of the depth of cognitive processing. The second way was that we wanted to use a different scale to measure the feelings of eeriness experienced by the participants. Mathur and Reichling (2016) managed to capture the uncanny valley curvature by presenting mechanistic faces ranging from very robot-like to faces with a high degree of human-likeness. In order to replicate their results we used half of their stimuli set; their stimuli were ranked from 1 to 80 and we used only the faces with uneven numbers in our experiment. Our results have shown that we were able to see and replicate the uncanny valley effect in the long condition. This is both in line with prior research and with our initial hypothesis. What is not in line with our hypothesis, however, is that the uncanny valley curvature is conspicuous by its absence in the short condition. Due to the high uncertainty and ambiguity of the data we cannot state with any reasonable accuracy that there is a valley in the short condition. Consequently, the hypothesis that the uncanny valley can be seen and replicated in a condition with short presentation times must be rejected. This was not an expected result because other hypotheses on the uncanny valley show that it is most likely explained by the fast and automatic processes (MacDorman et al., 2009). This was confirmed by Moll (2015) in his study on the effect of presentation time on ratings of uncanniness. Face processing is one of these fast processes, and studies on this have shown that a short presentation time is enough to accurately categorise faces even if their presentation time is less than 50 ms (Bar et al., 2006; Grill-Spector & Kanwisher, 2005; Stone et al., 2001). Therefore it is surprising that we were unable to clearly replicate the uncanny valley using short presentation times.

A secondary objective of our study was to compare the eeriness index of Ho and MacDorman (2010) to the scale used by Mathur and Reichling (2016). We found that the scales cannot be compared properly and are largely context-dependent because the difference in items is too large; the eeriness index has very extreme items, which brings the issue that when the stimuli are not perceived as those extremes, ratings will drift more towards the middle. Therefore the scale by Mathur and Reichling (2016) would have been more suitable in our study as we ended up not using half of our scale. Should the stimuli have been more extreme, it is likely that the eeriness index would have been the better choice. A pilot study could bring clarity on which scale to use. Ho and MacDorman (2010) did give the disclaimer that the index was developed and validated with a particular set of stimuli, and that it had yet to be tested using other sets.

The last objective of our study was to partly replicate a study by Moll (2015), who made use of morphed blends of human and robot faces. We wanted to determine if human reactions to android robots indeed exhibit the uncanny valley phenomenon, using that approach. By using artificial faces with varying degrees of human-likeness based on the morphing levels we expected to confirm the results of Moll (2015). In an attempt to do so, we used faces that were judged as somewhere in the middle of the mechano-humanness scale as estimated in the study by Mathur and Reichling (2016). This approach turned out to be erroneous because the morphing sequences we used ended up being too short; the mechanistic target faces were too close to the human faces. Because of this, we only see the upwards slope out of the valley and towards the genuine human face. The result is that we cannot estimate the position of the trough with reasonable accuracy and as such we cannot answer the research question satisfactory.

Category confusion

While the data in the short condition does not show the uncanny valley, it is does show that the estimated trough position makes a large shift to the left. This shift can be explained using category confusion, or category uncertainty. Contemporary literature explains this phenomenon as ambiguity that is experienced at the boundary between perceptual categories (Mathur & Reichling, 2016). This in turn could lead to trouble when it comes to determining the category to which an entity belongs (MacDorman & Chattopadhyay, 2016). A similar definition is given by Burleigh, Schoenherr, and Lacroix (2013) who state that category confusion occurs when stimuli that are in between non-human and human are perceived as ambiguous and consequently elicit negative emotions. The way category confusion is measured is as an increase in the time required categorising a stimulus (de Gelder, Teunisse, & Benson, 1997; Yamada, Kawabe, & Ihaya, 2013). While Mathur and Reichling (2016) merely suggest that this phenomenon may occur in the uncanny valley, other authors have gone so far as to propose that the valley of eeriness is even caused by category confusion (Burleigh et al., 2013; Green et al., 2008; Jentsch, 1906/1997; Yamada et al., 2013). Jentsch (1906/1997) for example developed a theory identifying category confusion as a cause of uncanniness. This was even before Mori's (1970) initial graph and hypothesis. Jentsch stated that feelings of eeriness were elicited by uncertainty about whether an entity is inanimate or animate, or whether it is nonhuman or human. Category confusion then occurs whenever an entity transitions from one category to the other, where the salience of the changes increases near the boundary of the categories. The result is an increased perception of eeriness for the observer. Another example supporting category confusion as a cause of the uncanny valley was found in a study by Burleigh et al. (2013). They confirmed their hypothesis stating that stimuli located in between two categories would elicit negative emotions due to conflicting representations.

New theoretical framework

We propose a new theoretical framework with category confusion as a basis. We hypothesise that there is a fast and early evaluation stating whether the observed stimulus is a human face or not a human face. Research has shown that a presentation time of 17 ms is enough to come to such a decision (Grill-Spector & Kanwisher, 2005; Mogg & Bradley, 1999; Stone et al., 2001). The next step in the process is the system firing off an answer after which an emotional response is experienced. This is followed up by a deeper inspection, and we propose that category confusion takes place somewhere between the initial categorisation and the deeper inspection. During deeper inspection conflicting information builds up as it is a cumulative process; the observer starts to notice the small differences that make the stimulus seem not as human-like as thought during the initial evaluation. The more salient the face is non-human, the faster the conflicting information builds up. The accumulated emotional response during deeper inspection only occurs if category confusion does take place. If there turns out to be no confusion on the category of the stimulus, the entire emotional asset of category confusion becomes non-existent because the participant proceeds to stick with their initial categorisation. So what happens if a stimulus is initially categorised as a human face even though it is clearly not? The category of said stimulus turns over, leading to negative judgment (Burleigh et al., 2013). The stronger the confusion, the more negative the response would then be. Figure 8 shows an illustration of our framework in the form of a flow chart.

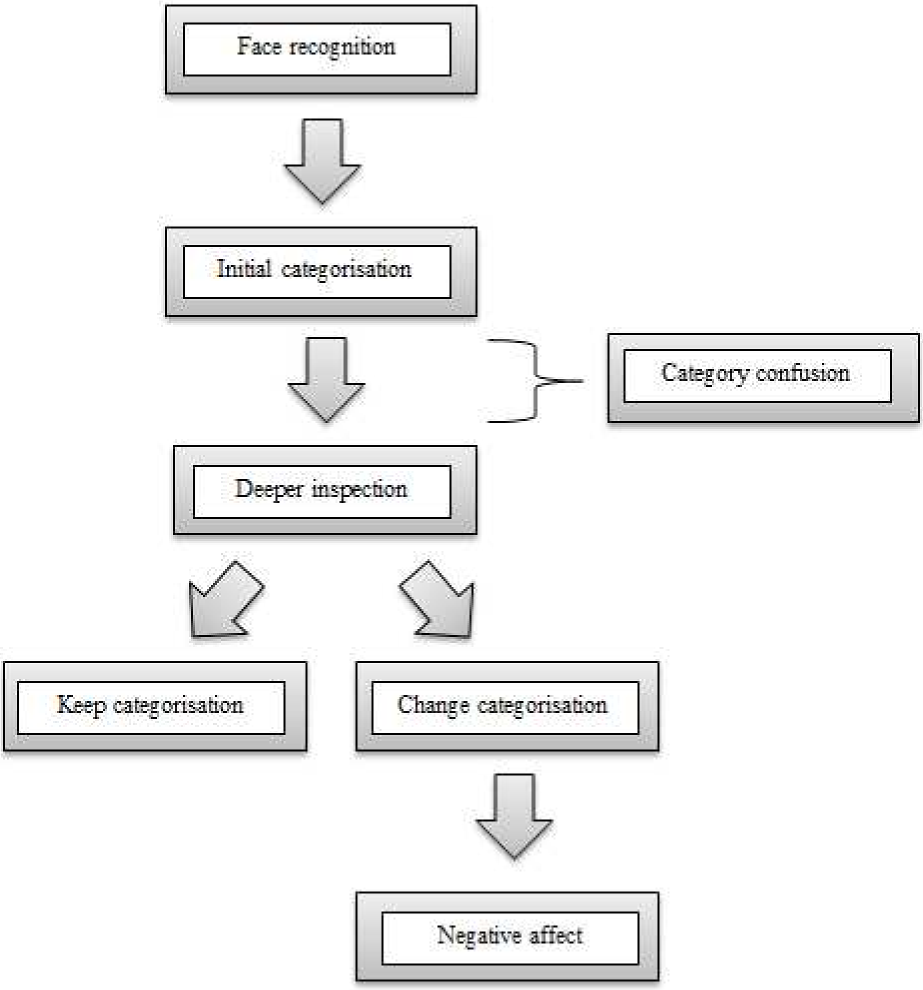

Figure 8. Flow chart of the low-level cognitive processes that we hypothesise to take place during perception of an artificial face

The important point is the category turnover, so if the process is cut off before the cumulative information has reached a critical point, the trough never happens. This can be done to find out at what point the trough is located, and if there is one at all, and as such is an interesting implication for future research. The critical point of information accumulation depends on the information that is given (i.e. the details of the stimulus). Consequently, the more time the deeper inspection takes, the more likely a change of category will take place.

Based on the workings of category confusion, we can explain the shift of the “trough” to the left because that is where we find the negative ratings that would be indicative of the uncanny valley. However, there are several issues with this reasoning. Firstly, based on the results we cannot even be sure that there is a trough; another way for interpretation is a monotonously rising curve. If it is no trough then we can also not be sure that there is a valley at all. Secondly, the absence of a clear uncanny valley could be a sign that the 50 ms used in our short condition is too little time for participants to categorise the stimuli. A third issue is that there is no consensus yet on whether category confusion really does cause the uncanny valley. For example, a recent study by MacDorman and Chattopadhyay (2016) showed no evidence to back up the claim that category confusion is a cause of the valley: the eeriest stimuli were categorised the fastest and with most certainty. Mathur and Reichling (2016) also state that "category confusion may occur in the UV but does not mediate the likability effect" (p. 22). Lastly, category confusion by itself cannot sufficiently explain the emotional aspect of the valley as there has been limited research on this. Yamada et al. (2013) conducted an experiment where intermediate morphs between various stages of human-likeness (real, hand-drawn, and stuffed-toy versions of human faces) elicited the longest categorisation latency as well as the lowest likeability.

Emotional response

At this point it is not clear yet how the emotional component is explained, or how the emotional response arises from category confusion. The theory of fluency processing (Winkielman, Schwarz, Fazendeiro, & Reber, 2003) supplements category confusion and is capable of elucidating this, however. This theory argues that an incongruence between what is expected and what is observed causes disfluent processing. Conversely, congruence leads to fluent processing. Fluency of processing establishes a link between category confusion and the emotional response. Many studies have shown or suggested that disfluent processing generally leads to negative emotions (e.g. Song, 2009; Winkielman et al., 2003). Fluent processing on the other hand leads to positive emotions (Kuchinke, Trapp, Jacobs, & Leder, 2009; Reber, Schwarz, & Winkielman, 2004; Schwarz & Clore, 2007; Song, 2009; Winkielman et al., 2003). To illustrate with a metaphor: imagine an olive. When one looks at an olive briefly, it is not improbable that the olive in question is mistaken for a grape. Now imagine eating the olive before realising that it is in fact an olive and not a grape; a first reaction would most likely be one of disgust, even though the person actually likes olives (see Yeomans, Chambers, Blumenthal, & Blake, 2008, for the role of expectancy in sensory evaluation). After all, there was incongruence between what is expected (a grape) and what is observed (an olive). Based on the workings of fluency processing we can state that this theory complements category confusion when it comes to explaining affect. We observed negative emotions and, using the theory of processing fluency, can then conclude that the participants experienced disfluent processing when observing the stimuli. While this does not correspond to the results of MacDorman and Chattopadhyay (2016), it is in line with the study by Yamada et al. (2013). Again there appears to be lack of a clear consensus, this time on whether negative emotions are caused by fluent or by disfluent processing. Furthermore, similar to category confusion it is possible that 50 ms is too short for proper fluency of either kind to take place. This seems to correspond with an article by Schwarz (2004), which mentions a number of relevant variables that can influence the speed and accuracy of processes concerned with identifying a stimulus' identity or form. Among these variables mentioned is the duration of its presentation, supporting the possibility of the presentation time being too short to fully process the stimulus and therefore leading to the absence of not only congruence, but also incongruence.

To conclude: many studies give category confusion as an explanation for the uncanny valley. Because we observe the ratings in our short condition as largely negative, it is plausible to use category confusion as a basis. Processing fluency is used complementary and explains the emotional response; the negative affect is caused by disfluent processing, or a discrepancy between what is expected and what is observed. Both theories are not without criticism, however, and for either or both to be applied by participants a presentation time of more than 50 ms is necessary.

Implications of the present study

It is important to note that there are many adaptations of the original uncanny valley graph. Figure 9 shows an adaptation of the uncanny valley in grey by Mathur and Reichling (2016) laid over Mori's (1970) original graph used as a frame of reference in this study. Some striking differences can be seen; the main one being that in the adaptation still and moving images have been combined. This could explain a second observation, namely that the trough is further to the left on the human-likeness scale as well as somewhat less deep. Furthermore, the actual valley is in between the moving and the still graphs in terms of width, which corresponds to the mix of both into one graph. This overlay is important because it shows that there are some significant differences between the various versions of the uncanny valley that are used in contemporary literature. This makes it difficult to draw unequivocal conclusions across studies, as these may vary based on trough position or stimulus type. These in turn differ based on what version of the valley is used for reference. This is an important factor to keep in mind, as it spans across all research on the uncanny valley and is an area with much room for improvement.

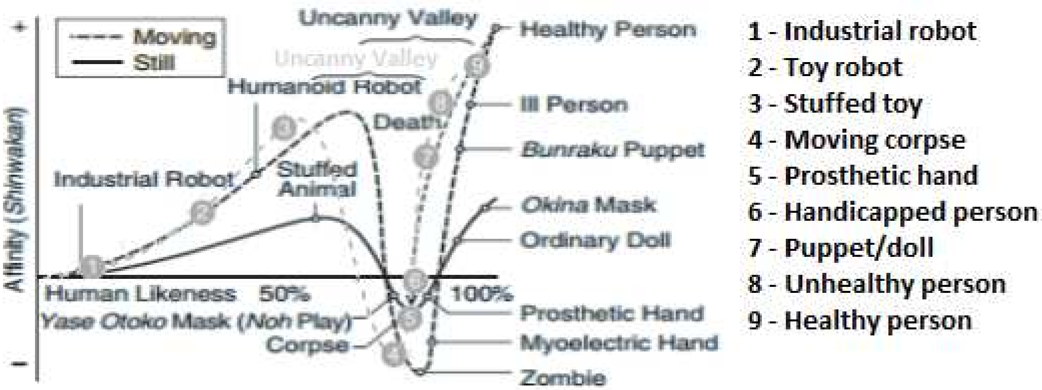

Figure 9. An adaptation of the uncanny valley where still and moving images are combined laid over the original graph

While not directly related to our research questions, an interesting finding to reflect upon is the position of the trough in our long condition. We notice that in our study, as well as in that of Mathur and Reichling's (2016), the trough is positioned surprisingly far to the left compared to the original depiction by Mori (1970). This means that the participants experienced the most eeriness when observing faces with a lower degree of human-likeness than hypothesised by Mori (1970). However, one must keep in mind that the stimuli used in both experiments were images of (humanoid) robots. In the original figure showing the valley there is a significant distinction made between still and moving images, with each having their own curvature. Still stimuli get a much less amplified peak and valley compared to moving stimuli (see Figure 10). Additionally, it is the curvature of moving stimuli that includes humanoid robots as one of the primary examples.

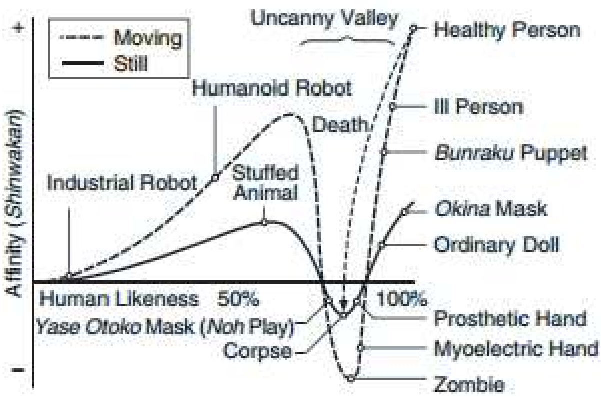

Figure 10. Original graph by Mori (1970) showing the difference in curvature between moving and still stimuli

So what, then, happens if you present the stimuli that are supposed to be moving according to the original graph as still? Many studies have opted or shown that moving stimuli increase human-likeness and, conversely, that still stimuli take away from human-likeness (e.g. Mori, 1970; Shimada & Ishiguro, 2008). The result is then that the entire valley makes a shift to the left when comparing to the original uncanny valley of moving stimuli.

For film producers and game artists, this study has shown that the ideal face is still equivocal; one should not adhere to a single graph of the uncanny valley. Because our results indicate that the valley may be located more to the left than initially thought, the climb out of the valley takes place at a lower level of human-likeness compared to other versions of the graph. This means that it may be easier to avoid feelings of eeriness for the observer than currently anticipated, as these artificial faces are easier to create than faces with 100% human-likeness. However, the opposite is also true: the descent and thereby negative affect is reached faster as well.

Limitations

There are several limitations to our study. The first of these is related to the quality of the morphs. Our morphs were merged in a way where the transition from one face to the other was made to be as smooth as possible, leading to an absence of extremely eccentric features. In retrospect this was a wrong way to create the morphing sequences; Seyama and Nagayama (2007) showed that bizarre facial features such as abnormal eyes increase the perception of the uncanny valley. These features were largely absent in our morphs, which could explain the fact we only captured the ascent out of the valley. The right upward slope leading out of it is clearly visible, rather than the valley in its entirety like we initially hypothesised. It seems that the closeness of the trough is not as near as we initially expected.

A second limitation is regarding the stimuli used for the replication of the study by Mathur and Reichling (2016). These faces showed massive variability in factors that could have influenced the way they were perceived. Proportions (Stirrat & Perret, 2010), positioning (Mara & Appel, 2015), background setting (Winkielman et al., 2003), sex (Bohnet & Zeckhauser, 2004; Buchan, Croson, & Solnik, 2008), resemblance to the viewer (DeBruine, 2002), and more, are all examples of these factors. Furthermore, the faces were all found using a search on the Internet and as such may be a biased representation of the total possible range of robots (Mathur & Reichling, 2016). Factors such as intended audiences could be confounding variables of the relationship between the human-likeness of the faces and the responses of the participants.

Thirdly, we observed a high uncertainty of our trough in comparison to Mathur and Reichling (2016). There is a significant influence of sample size on certainty; research has indicated that a low sample size can cause low statistical power (Button, 2014; Button et al., 2013; Fraley & Vazire, 2014). Low statistical power is defined as the probability that the null hypothesis is correctly rejected when it is false (Button et al., 2013). This means that when it is low in our study, there is also a low chance our hypothesis is correctly rejected and thus a high chance on a type II error. Furthermore, low statistical power reduces the ability to detect experimental effects (Krzywinski & Altman, 2013), and Button et al. (2013) show that a small sample size undermines the reliability of neuroscience, a field closely related to cognitive psychology (Eysenck & Keane, 2000). As the sample used in our study is not particularly large, this could explain the large degree of uncertainty in comparison to that of Mathur and Reichling (2016). They used many more participants and while each participant was shown only a small subset of stimuli, the total number of all observations was much higher (2800 versus 5130). This resulted in much more certainty of the results. Ways to improve upon our high uncertainty is through increasing the sample size. An a priori power analysis combined with many more stimuli per participant could have prevented the uncertainty of our trough estimates.

Future research

We propose an experiment where the participants have to perform a two-choice task. The same stimuli would be used as in this experiment but they would instead be presented in pairs. This diffusion model for two-choice decisions shows how stimulus information guides decisions and shows how the information is processed through time (Ratcliff & Rouder, 1998). The model assumes that decisions are made by a noisy process that accumulates information over time from a starting point towards one of two response criteria (Ratcliff & McKoon, 2008). Furthermore, the diffusion model as described is closely related to our assumption on the workings of category confusion and deeper inspection discussed earlier.

Within the same participant, two different things can happen: category confusion either takes place or it does not. The main measurements of the study would be the amount of confusion and conflict experienced by the participants, so it is key to look at the responses a particular stimulus gets.

Research should also be done on what point a participant is able to process the incongruence between what is expected and what is observed; based on our results it is possible that a window of 50 ms is not enough to do so. We therefore suggest conducting a similar study but instead using presentation times between 50 ms and 5 sec, narrowing down the window between congruence and incongruence and providing more insight on why we observed a shift to the left. These studies should not only focus on increasing the presentation time but also the variety: more than two conditions and with regular increments. This would allow for a more detailed estimate on the time that is needed for a participant to realise the observed stimulus is not in fact entirely human. As a result, a better understanding of which theories can be used to explain the uncanny valley may emerge and perhaps some of the current disagreement in contemporary literature will be resolved.

Furthermore, future studies should have the mechanistic target faces in the morphing sequences be less focused towards the human faces. In order to be able fully capture the entire uncanny valley curvature, the range of faces should not be as short as we have used in our study. Studies have indicated that it may help to ensure the target faces have sufficient eccentric features as well in order to bring out any feelings of eeriness in the participants.

Conclusion

This study attempted to replicate research on the uncanny valley; its goal was to confirm the existence of the valley in conditions with both short and long presentation times. In doing so, both images of robots as well as images of morphs between humans and robots with a high degree of human-likeness were used. The results have shown that there is indeed a clear uncanny valley noticeable when participants have a long time to observe the robotic faces. However, they also showed that there is no clear valley when participants only have a very brief moment before the stimulus disappears. Possible explanations for its conspicuous absence are related to category confusion and fluency of processing, but for either to be applied a longer presentation time may be necessary. Furthermore, we found that the eeriness index (Ho and MacDorman, 2010) is mainly suitable when the stimuli are near the extremes of the scale. A last conclusion is that when using morphed faces, it is imperative to use robot faces that range from one end of the scale to another. Not doing so causes the morphing sequences to be too short, resulting in capturing only one side of the uncanny valley. Future research should incorporate the diffusion model into the experiment and aim to increase presentation times and increments, use faces that properly cover the robot to human-likeness spectrum, and increase the sample size. Doing so should show with more certainty whether the uncanny valley occurs with short presentation times as well. If not, does the valley disappear gradually or is there a sudden threshold that would indicate a required level of cognitive processing?

Acknowledgements

The author would like to thank the first supervisor Dr. M. Schmettow for his patience, for providing excellent guidance and insight where necessary, and for his assistance with the statistical analyses. Further thanks go out to Dr. M. L. Noordzij for his helpful advice and comments. Lastly, thanks to B. Moll, M. Mathur, and D. Reichling for the transparency they provided of their data, allowing us to attempt a replication of their research.

References

Bailenson, J. N., & Yee, N. (2005). Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science, 16*(10), 814-819. doi: 10.1111/j.1467-9280.2005.01619.x

Boker, S. M., Cohn, J. F., Theobald, B. J., Matthews, I., Mangini, M., Spies, J. R., … & Brick, T. R. (2011). Something in the way we move: Motion dynamics, not perceived sex, influence head movements in conversation. *Journal of Experimental Psychology: Human Perception and Performance, 37*(3), 874-891. doi: 10.1037/a0021928

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6*(2), 269-278. doi:10.1037/1528-3542.6.2.269

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics, 1*(1), 71-81. doi:10.1007/s12369-008-0001-3

Bemelmans, R., Gelderblom, G., Jonker, P., & de Witte, L. (2012). Socially assistive robots in elderly care: A systematic review into effects and effectiveness. *Journal of the American Medical Directors Association, 13*(2), 114-120. doi:10.1016/j.jamda.2010.10.002

Bohnet, I., & Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior and Organization, 55*(4), 467-484. doi: 10.1016/j.jebo.2003.11.004

Buchan, N. R., Croson, R. T. A., & Solnik, S. (2008). Trust and gender: An examination of behavior and beliefs in the investment game. *Journal of Economic Behavior and Organization, 68(3-4)*, 466-476. doi: 10.1016/j.jebo.2007.10.006

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

Butler, M., & Joschko, L. (2009). Final Fantasy of The Incredibles: Ultra-realistic animation, aesthetic engagement and the uncanny valley. *Animation Studies, 4*, 55-63.

Button, K. S. (2014). Unreliable neuroscience? Why power matters. Retrieved from https://www.theguardian.com/science/sifting-the-evidence/2013/apr/10/unreliable-neuroscience-power-matters

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365-376. doi: 10.1038/nrn3475

Canemaker, J. (2004). *A part-human, part-cartoon species*. The New York Times. Retrieved from http://www.nytimes.com/2004/10/03/movies/a-parthuman-partcartoon-species.html?_r=0

Chaminade, T. D., Hodgins, J. K., & Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters' actions. *Social Cognitive and Affective Neuroscience, 2*(3), 206-216. doi:10.1093/scan/nsm017

Cheetham, M., Suter, P., & Jäncke, L. (2011). The human likeness dimension of the "Uncanny Valley Hypothesis": Behavioral and functional MRI findings. *Frontiers in Human Neuroscience, 5,* 1-14. doi:10.3389/fnhum.2011.00126

DeBruine, L. M. (2002). Facial resemblance enhances trust. *Proceedings of the Royal Society B: Biological Sciences, 269*(1498), 1307-1312. doi: 10.1098/rspb.2002.2034 de Gelder, B., Teunisse, J. P., & Benson, P. J. (1997). Perception of facial expressions: Categories and their internal structure. *Cognition and Emotion, 11*(1), 1-23. doi: 10.1080/026999397380005

European Conference on Visual Perception. (2008). Retrieved from pics.psych.stir.ac.uk

Eysenck, M. W., & Keane, M. T. (2000). Cognitive psychology: A student's handbook (4th Ed.). Hove and London, UK: Lawrence Erlbaum Associates.

Entertainment Software Association. (2015). *Essential facts about the computer and video game industry*. Retrieved from http://www.theesa.com/wp-content/uploads/2015/04/ESA-Essential-Facts-2015.pdf

Fraley, R. C., & Vazire, S. (2014). The N-pact factor: Evaluating the quality of empirical journals with respect to sample size and statistical power. *PLoS ONE, 9*(10), e109019. doi: 10.1371/journal.pone.0109019

Freedman, Y. (2012). Is it real… or is it motion capture? The battle to redefine animation in the age of digital performance. *The Velvet Light Trap, 69*, 38-49. doi: 10.1353/vlt.2012.0001

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition, 125*(1)*,* 125-130. doi:10.1016/j.cognition.2012.06.007

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science, 16*(2), 152-160. doi:10.1111/j.0956-7976.2005.00796.x

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

Ho, C., MacDorman, K. F., & Pramono, Z. A. D. (2008). Human emotion and the uncanny valley: A GLM, MDS, and Isomap analysis of robot video ratings. In T. Fong and K. Dautenhahn (Eds.), *Proceedings of the Third ACM/IEEE International Conference on Human-Robot Interaction* (pp. 169-176). New York, NY: ACM. doi:10.1145/1823738.1823740

Iromec (2009). *Interactive robotic social mediators as companions.* Retrieved from http://www.Iromec.org

Jentsch, E. (1906/1997). On the psychology of the uncanny (R. Sellars, Trans.). *Angelaki, 2*(1), 7-16. doi: 10.1027/1864-1105/a000156

Joly, J. (2008). *Can a polygon make you cry?* Retrieved from https://jonathanjoly.wordpress.com/

Kaba, F. (2013). Hyper-realistic characters and the existence of the uncanny valley in animation films. *International Review of Social Sciences and Humanities, 4*(2), 188-195.

Keysers, C., & Gazzola, V. (2007). Integrating simulation and theory of mind: From self to social cognition. *Trends in Cognitive Sciences, 11*(5), 194-196. doi:10.1016/j.tics.2007.02.002

Kolers, P. A. (1968). *Some psychological aspects of pattern recognition.* In P. A. Kolers and M. Eden (Eds.)*, Recognizing patterns* (pp. 4-61). Cambridge, Massachusetts: MIT Press.

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fmri. *PloS One, 3*(7), e2597. doi:10.1371/journal.pone.0002597

Krzywinski, M., & Altman, N. (2013). Points of significance: Power and sample size. *Nature Methods, 13*(12), 1139-1140. doi: 10.1038/nmeth.2738

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallarm, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin, 126*(3), 390-423. doi:10.1037/0033-2909.126.3.390

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

MacDorman, K. F., & Entezari, S. O. (2015). Individual differences predict sensitivity to the uncanny valley. *Interaction Studies, 16*(2), 141-172. doi:10.1075/is.16.2.01mac

MacDorman, K. F., Green, R. D., Ho, C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior, 25*(3), 695-710. doi: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. doi:10.1075/is.7.3.03mac

MacDorman, K. F., Minato, T., Shimada, M., Itakura, S., Cowley, S., & Ishiguro, H. (2005). Assessing human likeness by eye contact in an android testbed. In B. G. Bara, L. Barsalou, and M. Bucciarelli (Eds.), *Proceedings of the XXVII Annual Meeting of the Cognitive Science Society* (pp. 1373-1378). Stresa, Italy.

MacDorman, K. F., Vasudevan, S. K., & Ho, C. (2008). Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *AI & Society, 23*(4), 485-510. doi: 10.1007/s00146-008-0181-2

MacMillan, D. (2007). *Navigating the uncanny valley.* Business Week. Retrieved from http://businessweek.com/innovate/content/aug2007/id20070817_955317.htm

Mara, M., & Appel, M. (2015). Effects of lateral head tilt on user perceptions of humanoid and android robots. *Computers in Human Behavior, 44()*, 326-334. doi: 10.1016/j.chb.2014.09.025

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

Matsui, D., Minato, T., MacDorman, K. F., & Ishiguro, H. (2005). Generating natural motion in an android by mapping human motion. In M. Meng and H. Zhang (Eds.), *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 3301-3308). Edmonton, Canada: IEEE.

Minato, T., Shimada, M., Ishiguro, H., & Itakura, S. (2004). Development of an android robot for studying human-robot interaction. In R. Orchard, C. Yang, and M. Ali (Eds.), *Innovations in applied artificial intelligence* (pp. 424-434). Berlin, Germany: Springer-Verlag.

Mogg, K., & Bradley, B. P. (1999). Orienting of attention to threatening facial expressions presented under conditions of restricted awareness. *Cognition and Emotion, 13*(6), 713-740. doi:10.1080/026999399379050

Moll, B. (2015). *Investigating the origins of the uncanny valley: The effect of presentation time on ratings of uncanniness* (Unpublished master thesis). University of Twente, Enschede, the Netherlands.

Mori, M. (1970/2012). The uncanny valley (K. F. MacDorman & N. Kageki, Trans.). *IEEE Robotics & Automation Magazine, 19*(2), 98-100. doi: 10.1109/MRA.2012.2192811

Öhman, A. (2000). Fear and anxiety: Evolutionary, cognitive, and clinical perspectives. In M. Lewis and J. M. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 573-593). New York, NY: The Guilford Press.

Olson, I. R., & Marshuetz, C. (2005). Facial attractiveness is appraised in a glance. *Emotion, 5*(4), 498-502. doi:10.1037/1528-3542.5.4.498

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716-aac4716. doi: 10.1126/science.aac4716

Peca, A., Simut, R., Pintea, S., Costescu, C., & Vanderborght, B. (2014). How do typically developing children and children with autism perceive different social robots? *Computers in Human Behavior, 41,* 268-277. doi:10.1016/j.chb.2014.09.035

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*(4), 873-922. doi: 10.1162/neco.2008.12-06-420

Ratcliff, R., & Rouder, J. N. (1998). Modelling response time for two-choice decisions. *Psychological Science, 9*(5), 347-356. doi: 10.1111/1467-9280.00067

Reips, U. D., & Funke, F. (2008). Interval-level measurement with visual analogue scales in Internet-based research: VAS generator. *Behavior Research Methods, 40*(3), 699-704. doi: 10.3758/BRM.40.3.699

Rozin, P., & Fallon, A. E. (1987). A perspective on disgust. *Psychological Review, 94*(1), 23-41. doi:10.1037/0033-295X.94.1.23

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2011). The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience, 8*(4), 413. doi: 10.1093/scan/nsr025

Schwarz, N. (2004). Meta-cognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology, 14*(4), 332-348. doi: 10.1177/0272989X04273144

Schwarz, N., & Clore, G. L. (2007). Feelings and phenomenal experiences. In A. Kurglanski & E. T. Higgins (Eds.), *Social Psychology. Handbook of basic principles* (2nd ed., pp. 385-407). New York: Guilford.

Seyama, J., & Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments, 16*(4), 337-351. doi:10.1162/pres.16.4.337

Shimada, M., & Ishiguro, H. (2008). Motion behavior and its influence on human-likeness in an android robot. In V. Sloutsky, B. C. Love, & K. McRae (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2468-2473). Washington, DC: Cognitive Science Society.

Soler, C., Núñez, M., Gutiérrez, R., Núñez, J., Medina, P., Sancho, M. … Núñez, A. (2003). Facial attractiveness in men provides clue to semen quality. *Evolution and Human Behavior, 24*(3), 199-207. doi:10.1016/S1090-5138(03)00013-8

Song, H. (2009). *The effects of processing fluency on judgment and processing style: Three essays on effort prediction, risk perception, and distortion detection.* Unpublished doctoral dissertation, The University of Michigan, USA.

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science, 21*(3), 349-354. doi: 10.1177/0956797610362647

Stone, A., Valentine, T., & Davis, R. (2001). Face recognition and emotional valence: Processing without awareness by neurologically intact participants does not simulate covert recognition in prosopagnosia. *Cognitive, Affective, & Behavioral Neuroscience, 1*(2), 183-191. doi:10.3758/CABN.1.2.183

Tapus, A., Ţăpuş, C., & Matarić, M. J. (2008). User-robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intelligent Service Robotics, 1*(2), 169-183. doi:10.1007/s11370-008-0017-4

Tinwell, A., Nabi, D. A., & Charlton, J. P. (2013). Perception of psychopathy and the Uncanny Valley in virtual characters. *Computers in Human Behavior, 29*(4)*,* 1617-1625. doi:10.1016/j.chb.2013.01.008

Tinwell, A., Nabi, D. A., Grimshaw, M., & Williams, A. (2011). Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Computers in Human Behavior, 27*(2), 741-749. doi:10.1016/j.chb.2010.10.018

Travers, P. (2001). Final Fantasy. Retrieved from http://www.rollingstone.com/movies/reviews/final-fantasy-20010706

Von der Pütten, A. M., Krämer, N. C., Gratch, J., & Kang, S. (2010). "It doesn't matter what you are!" Explaining social effects of agents and avatars. *Computers in Human Behavior, 26*(6), 1641-1650. doi: 10.1016/j.chb.2012.06.012

Wada, K., Shibata, T., Asada, T., & Musha, T. (2007). Robot therapy for prevention of dementia at home – Results of preliminary experiment. *Journal of Robotics and Mechatronics, 19*(6)*,* 691-697. doi: 10.20965/jrm.2007.p0691

Willis, J. & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592-598. doi:10.1111/j.1467-9280.2006.01750.x

Winkielman, P., Schwarz, N., Fazendeiro, T., & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch & K. C. Klauer

(Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 189-217). Mahwah, NJ: Lawrence Erlbaum.

Yamada, Y., Kawabe, T., & Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the “uncanny valley” phenomenon. *Japanese Psychological Research, 55*(1), 20-32. doi: 10.1111/j.1468-5884-2012-00538.x

Yeomans, M. R., Chambers, L., Blumenthal, H., & Blake, A. (2008). The role of expectancy in sensory and hedonic evaluation: The case of smoked salmon ice-cream. *Food Quality and Preference, 19*(6), 565-573. doi:10.1016/j.foodqual.2008.02.009

Appendices

Appendix A

Morphing percentage per sequence	Morph 1	Morph 2	Morph 3	Morph 4	Morph 5	Morph 6	Morph 7	Morph 8	Morph 9	Morph 10
Sequence A	0%	16%	33%	49%	65%	72%	79%	86%	93%	100%
Sequence B	0%	13%	26%	39%	52%	62%	71%	81%	90%	100%
Sequence C	0%	14%	28%	42%	57%	66%	74%	83%	91%	100%
Sequence D	0%	15%	30%	45%	60%	68%	76%	84%	92%	100%

Appendix B

English items		Dutch items		German items	
Anchor low	Anchor high	Anchor low	Anchor high	Anchor low	Anchor high
Reassuring	Eerie	Geruststellend	Griezelig	Beruhigend	Gruselig
Numbing	Feaky	Uitdrukkingsloos	Eng	Ausdruckslos	Unheimlich
Ordinary	Supernatural	Gewoontjes	Bovennatuurlijk	Gewöhnlich	Übernatürlich
Uninspiring	Spine-tingling	Zonder enthousiasme	Opwindend	Wenig begeisternd	Aufregend
Boring	Shocking	Saai	Schokkend	Langweilig	Schockierend
Predictable	Thrilling	Voorspelbaar	Spannend	Vorhersehbar	Spannend
Bland	Uncanny	Nietszeggend	Verontrustend	Nichtssagend	Beunruhigend
Unemotional	Hair-raising	Emotieloos	Doodeng	Emotionslos	Haarsträubend

Appendix C

```r
D_[["MathurRepl_3"]] <-
  D_[["MathurRepl"]] %>%
  filter(Collection == "Mathur") %>%
  mutate(Condition = str_c(Experiment, "_", Condition)) %>%
  group_by(Condition, Stimulus, huMech) %>%
  summarize(avg_response = mean(response)) %>%
  mutate(huMech0 = 1,
         huMech2 = huMech^2,
          huMech3 = huMech^3) %>%
  rename(huMech1 = huMech)


F_[["MathurRepl_3"]] <-
  formula("avg_response ~ 0 + (huMech0 + huMech1 + huMech2 + huMech3):Condition")


# D_[["MathurRepl_3"]] %>%
# lm(F_[["MathurRepl_3"]], data = .)


M_[["MathurRepl_3"]] <-
  D_[["MathurRepl_3"]] %>%
  brm(F_[["MathurRepl_3"]],
      data = .,
      chains = 3)


## extracting posterior and fitted values
P_[["MathurRepl_3"]] <-
  tbl_post(M_[["MathurRepl_3"]])


# save(D_, P_, M_, T_, F_, C_, G_, file = "DK.Rda")
```