Taking a step back to see more clearly: Exploring the full extent of the Uncanny Valley using biological faces

Bachelor Thesis – Milan Bischoff
University of Twente, 2021

1st Supervisor: Dr. Martin Schmettow
2nd Supervisor: Simone Borsci

Abstract

Despite two decades of research, the causes of the Uncanny Valley remain a mystery. Roboticist Masahiro Mori proposed the effect more than 50 years ago to describe the strongly negative and aversive reaction experienced in response towards artifacts of high, yet imperfect human resemblance. For the most part, the increasing human resemblance of a robot makes it more likeable, yet Mori correctly predicted the sudden drop of affinity in response to near realistic-looking robots. However, just like the evaluation of the perceived emotional response, the only known and reliable method to determine the human-likeness of a robot is through implicit ratings which leave the independent variable and subsequently, the factors responsible for the feeling ignored and unexplored. To shed some light on the possible conceptualizations of human-likeness and in regard to the prominence of evolutionary explanations to the Uncanny Valley effect we exploratively tried to replicate the effect outside of its original domain of engineering. Consequently, we distanced ourselves from the usual experimental setup using artificial faces of robots or computer-generated characters and instead created a stimulus set with only unmanipulated biological faces of human and non-human primates. Additionally, as an alternative and objective measurement of human-likeness the ancestral closeness of presented primates to the homo sapiens was added to the analysis. Nonetheless, the main analysis was conducted with averaged subjective ratings of the human-likeness by the research team. The emotional reaction of participants towards the primate faces was measured by an eeriness index displayed on a visual analogue scale. Using multilevel modelling we were able to observe the effect almost universally throughout all participants individually. However, the Uncanny Valley did not show at the population-level using the averaged responses of participants. Furthermore, a comparison of the individual onset of the effect of all participants revealed that the Uncanny Valley is elicited consistently at the same level of human-likeness. Lastly, we found that the ancestral closeness of primates was congruent to the ratings of human-likeness, showing that the phylogenetic similarity of primates to the homo sapiens can also be successfully employed as a measure of human-likeness. Overall, these results allow us to generalize the Uncanny Valley effect as a broader phenomenon independent of artificial or synthetic characters. This strongly favours the evolutionary approach stating that the Uncanny Valley is a particular manifestation of a mechanism that originally developed to increase our reproductive fitness over the course of evolution. Based on this approach the feeling of aversion might have originally served us to prevent reproduction with unfit individuals and other species or to avoid contracting transmittable diseases from closely related species.

Keywords: Uncanny Valley, primates, evolutionary perspective, visual perception, phylogenetics

1. Introduction

1.1 Motivation

Today's robots must be likeable in addition to their functionality given their use in social settings such as healthcare, education, entertainment or even as museum guides is well under its way (McTear, Callejas, & Griol, 2016; Dietsch, 2010). To ease the interaction, especially for people unfamiliar with new technology, robots have been designed physically resemblant to humans (Kätsyri, Förger, Mäkäräinen, & Takala, 2015; Mathur & Reichling, 2016; Mori et al., 1970/2012). However, in strong contrast to the intention and unlike their mechanistic counterparts, humanoid robots such as the recreation of sci-fi writer Philip K. Dick or 'Sophia the Robot' from Hanson Robotics frequently cause frightening reactions and leave their observers perturbed (Hanson et al., 2005; AgoraTec, 2018). The same aversive reaction has been observed with computer generated (CG) faces in movies or videogames. Animators behind the DreamWorks movie 'Shrek' or the videogame 'Final Fantasy: The spirits within' have pointed out "distinctly unpleasant" or "grotesque" sensations respectively when their designed characters began looking too human-like but lacked the level of realism to quite convince their viewers (Brenton, Gillies, Ballin, & Chatting, 2005).

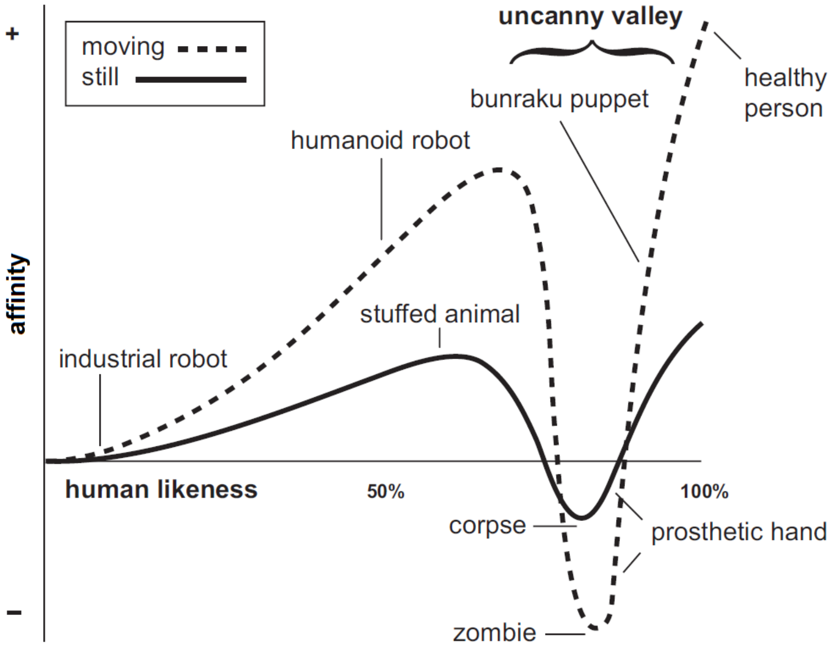


Fig. 1 The Uncanny Valley model adapted from MacDorman and Ishiguro (2006)

Engineers, as well as animators, are long aware of this phenomenon and the threat it presents to their profession due to Masahiro Mori's foresighted observation. More than 50 years ago he coined the effect the 'Uncanny Valley' (see Fig. 1) describing a steep drop in people's affinity towards artifacts of high, yet imperfect human resemblance (Mori et al., 1970/2012). More precisely, the slow upward trend in likeability that comes with increasing human-likeness is disrupted by the 'valley' which represents the plummeting and recovering emotional reaction at medium to high human-likeness. Different descriptions arose over time to characterize the emotional response but most commonly it is referred to as an intense feeling of eeriness, strangeness, or unease (MacDorman, Green, Ho, & Koch, 2009; Brenton, Gillies, Ballin, & Chatting, 2005; Ho & MacDorman, 2017; Zhang et al., 2020). Mori et al. (1970/2012) initially observed the effect in relation to prosthetic arms, dolls, or toy robots. However, he quickly picked up on the implications on his domain of expertise – robotics. In the present study, we aim to generalize the current conception of the effect by trying to replicate it using only biological faces.

1.2 Contemporary research

By today a considerable number of articles have confirmed the theory of the Uncanny Valley and some even replicated his hypothesized Valley-model with data (Mathur et al., 2020; Slijkhuis & Schmettow, 2017; Keeris & Schmettow, 2016; MacDorman, Green, Ho, & Koch, 2009; Seyama & Nagayama, 2007; Haeske & Schmettow, 2016; Burleigh, Schoenherr, & Lacroix, 2013). Despite the increasing agreement on the existence of the Uncanny Valley contemporary research on the topic fails to provide sufficient evidence to outline what factors trigger the effect. Consequently, recent articles have suggested a broader stance towards the topic. In contrast to its original conception and domain, they proposed the Uncanny Valley not to be a mere engineering problem but rather a specific and observable facet of a more general psychological phenomenon (Moore, 2012; Mathur & Reichling, 2016). This notion is reflected by the dominance of psychological theories embedded in the explanatory framework to the phenomenon.

Contemporary theories explain the Uncanny phenomenon from either cultural, evolutionary, and purely cognitive perspectives or as a result of individual differences (Wang, Lilienfeld, & Rochat, 2015; Hanson et al., 2005; Zhang et al., 2020; Brenton, Gillies, Ballin, & Chatting, 2005). To distinguish the variety of theories they are separated into two categories. On the one hand there are fast-, automatic cognitive processes whereas on the other hand we find slow and conscious cognitive processes (MacDorman, Green, Ho, & Koch, 2009; Haeske & Schmettow, 2016). While evolutionary explanations fall into the first category, cultural as well as individual explanations belonging to the second (Haeske & Schmettow, 2016; Wang, Lilienfeld, & Rochat, 2015). The difference lies within the more complex, conscious and thus time-intensive reflection of personal attitudes and norms required by the cultural and individual explanations whereas the evolutionary explanations suggest specialized, implicit, and stimulus-driven processes that became hard-wired in our perception as a result of natural selection (Wang, Lilienfeld, & Rochat, 2015).

Recent research has provided evidence to prefer the fast-system theories and consequent evolutionary origin. Most notably, Slijkhuis and Schmettow (2017) showed the Uncanny Valley could be consistently replicated with presentation times as low as 100ms. Similarly, Haekse (2016) showed that the ratings of eeriness after 100ms of presentation time significantly predicted eeriness ratings based on unlimited presentation times. This confirms that the effect is caused predominantly by fast and specialized processes. Furthermore, Slijkhuis and Schmettow (2017) controlled the involvement of slow processing by presenting a mask immediately after the disappearance of stimuli that interrupted processing. Thus, it can be concluded that the effect can also occur independent of slower, more conscious cognitive processes (Slijkhuis & Schmettow, 2017). Additionally, Koopman and Schmettow (2019) outlined the effect to be universal by using a multilevel analysis to reveal that the Uncanny curvature was observable for each of their participants. Showing the universality of the effect within humans already provides strong ground to argue that the effect must have developed at a much earlier stage of human evolution or else it would not occur consistently in all humans. However, Steckenfinger and Ghazanfar (2009) further strengthened this conclusion by replicating the same effect within monkeys. In a creative research design, they used monkeys as their research subject and gaze duration as a measurement for the emotional reaction.

Yet, despite the reliance on psychological theories to provide a framework for the effect and the growing support towards an evolutionary explanation, contemporary studies almost exclusively used artificial faces of either real robots or morphed pictures of human and robot faces to test the Uncanny Valley. The absence of other types of stimuli might be explained by the lack of attention paid to the conceptualization of human-likeness (Zhang et al., 2020; Wang, Lilienfeld, & Rochat, 2015). Much like the participants' responses on the perceived eeriness of stimuli, the dimension of human-likeness is constructed exclusively based on implicit ratings of the perceived human-likeness of stimuli to circumvent the issue that the concept has no clear definition (Zhang et al., 2020). However, Rádlová, Landová, and Frynta (2018) showed how assessing a broader range of stimulus types yields evidence for a more well-founded reasoning on what factors make a face likeable or send it down the Uncanny Valley. In their study examining human attractiveness ratings of primate faces they found morphometrical variance to human faces to be a significant predictor for attractiveness ratings yet only when observing primates closely related to modern humans. Meanwhile the variance in face proportions was uncorrelated to ratings of more distant human relatives. Thus, neglecting the definition of human-likeness presents a serious problem. The absence of clear-cut criteria to determine a stimulus' level of human-likeness as well as the lack of a holistic approach prevents us to say with certainty why specific stimuli at a given level of human-likeness elicit the Uncanny Valley effect.

With regard to this lack, the current study pursues a more exploratory approach. It takes a much broader stance towards the Uncanny Valley and aims to shed more light on its potential evolutionary origin by exploring the occurrence of the effect outside of its usual sphere of robotic and CG faces. To do so the study design will include only biological and unmanipulated faces to either support or speak against the claim that a more general, evolutionary mechanism is being at work behind the Uncanny Valley.

1.3 Explanations of the Uncanny Phenomenon

As already introduced, the abundance of theoretical explanations on the Uncanny Valley is summarized within two categories (MacDorman, Green, Ho, & Koch, 2009; Haeske & Schmettow, 2016; Zhang et al., 2020). In the following the first category of automatic, stimulus-driven, and specialized perceptual processes will be referred to as fast system theories. In contrast, the second category including the more complex cognitive processes and conflicts is labelled as the slow system theories. However, the common premise of both categories is the sensation of conflicting cues (Wang, Lilienfeld, & Rochat, 2015). Therefore, we will outline the contemporary understanding of human face perception and explain what influences it underlies prior to elaborating on the categories and their respective theories. Firstly, this helps to differentiate the categories from another by highlighting how the fast system theories occur at a temporally distinct processing stage than the slow system theories. Moreover, it provides a framework of the hierarchy of cognitive processes in visual perception that provides a scientific underpinning to the core concepts of the later mentioned theories. However, last and most importantly, it underscores how the Uncanny Valley can be understood as a general problem of perceptual dissonance. further supports the notion that the effect is unlikely to be limited only to the domain of engineering.

1.3.1 Face perception

Face recognition holds distinct evolutionary significance based on its importance for successful social interaction (Zhao et al., 2018; Zhao, Chellappa, Phillips, & Rosenfeld, 2003). Consequently, the human face is proclaimed to be the most distinctive part of our body and most substantial for our ability to interact with one another (Yu, 2001). Due to this, most research on the Uncanny Valley has used the faces as their stimuli (Zhang et al., 2020). It should furthermore come as no surprise that face perception is a very intricate and highly specialized process. Thus, to effectively unravel the Uncanny phenomenon in its entirety we must understand this intricacy.

A multitude of articles has pointed out how high levels of human-likeness and increased realism of humanoid robots caused observers to be less tolerant towards imperfections within the face of the robot (Brenton, Gillies, Ballin, & Chatting, 2005; MacDorman & Entezari, 2015; Green, MacDorman, Ho, & Vasudevan, 2008). Similarly, higher sensitivity towards and stricter evaluation of the faces was also reported for increasingly human-like primates (Rádlová, Landová, & Frynta, 2018). This increase in sensitivity is also reflected in a higher discrimination ability for conspecifics than other species (Pascalis et al., 2005; Wang, Lilienfeld, & Rochat, 2015). Recently, Papeo (2020) provided an excellent explanation for this by showing how our visual perception is primed towards the detection of social interaction allowing us to recruit more attention for situations we perceived to be of social character in opposition to non-social situations. In their study, they presented participants with a variety of drawings of two human bodies which either faced each other or stood with their backs to each other (Appendix A). Simultaneously conducted fMRI scans revealed that participants exhibited an increased brain activity for the facing bodies in comparison to the non-facing bodies. Through his study Papeo (2020) revealed an innate human bias towards the perception of faces and thus showed how social environments influence our perception. On basis of this we can propose an explanation to the above-mentioned articles, arguing that the increasing human-likeness of an entity, regardless of artificial or biological nature, draws upon more cognitive resources resulting in a more scrutinous evaluation.

Furthermore, Papeo (2020) found that facing bodies elicited more internal expectations about the visual environment in participants than non-facing bodies. This is congruent with the results of Currie and Little (2009) showing that human faces are of utmost importance for judgements on bodily and physical attractiveness. Their observation suggests that general assumptions on other humans are already derived solely upon the perception of faces. In unison, these two studies suggest that perceiving the possibility of social interaction through observing either ‘facingness’ as described by Papeo (2020), or the detection of a face already induces implicit hypotheses. More generally, this shows the involvement of top-down processing which we will outline in the following paragraph (Bruce & Young, 1986).

To improve our understanding of the intricate workings of human face recognition Bruce and Young (1986) proposed a hierarchical model of separate functional components. Almost 20 years later Grill-Spector and Malach (2004) published an extensive review of a full decade of fMRI studies reflecting how the proposed hierarchical model of Bruce and Young (1986) is manifested in the spatial organization of the human visual cortex. First and foremost, they confirmed that there are distinct areas within our visual cortex of which some activate more frequently and some less. The difference in activity corresponds to the function of the respective area meaning that areas with general functions activate to almost any visual task whereas more specialized areas respond more infrequently and only to more complex visual tasks. Moreover, Bruce and Young (1986) proposed in their original model that more specialized brain areas are susceptible to repetition and can adapt with frequent exposure. Interestingly, this plasticity of more complex brain areas was also verified by Grill-Spector and Malach (2004) who observed higher levels of repetition-suppression for certain stimuli within the more specialized brain areas. This repetition-suppression is referred to as the perceptual magnet (Feldman, Griffiths, & Morgan, 2009). In their article Feldman, Griffiths, and Morgan (2009) proposed it to be a mechanism of optimal statistical inference, pulling stimuli within a category towards the category prototype and causing within-group equivalence while subsequently increasing discrimination ability at the category boundary where the pulls of competing category prototypes negate each other. Important to understand is how learned categories (implicit as well as explicit) and their respective prototypes constitute our perception thus showing the influence of top-down processing on our basic perception (Feldman, Griffiths, & Morgan, 2009; Schyns & Oliva, 1999).

This is further supported by Grill-Spector and Malach (2004) who additionally observed the transition from bottom-up processing to top-down processing in the spatial organization of the human visual cortex. Basically, the areas of the visual cortex are organized along two orthogonal axes. The first axis corresponds to the hierarchical processing. Thus, brain areas along this axis are activated successively with areas located farther on the axis activating later and only in response to more complex visual stimuli such as faces while areas at the beginning are consistently active. Whereas the first axis represents the hierarchical organization, the second axis corresponds to functional specialization. It is positioned orthogonally at the high-complexity end of the first axis and entails functionally specialized brain areas. Due to the orthogonality of both axes, the brain areas along the second axis can respond simultaneous to signals at the end of the hierarchical axis. While the hierarchical axis corresponds to bottom-up processing, the second axis shows how top-down processes flexibly influence our perception once a certain degree of categorization is achieved. This seems to be the focal point where our perception is confronted with and influenced by known concepts and consequent hypotheses. Understanding the functioning of the hierarchical axis allows understanding face detection and face identification as temporally distinct processes while processes such as the identification of faces and the processing of perceived emotions can occur simultaneously along the second axis. As observed by Or and Wilson (2010) the identification of faces takes on average 31ms longer than the detection of faces if evaluated by a threshold of 75% discrimination accuracy. Thus, the detection of faces is likely situated at an earlier stage of the hierarchy than the full identification of a face. In contrast, Eimer and Holmes (2002) observed the identification of faces and their displayed emotion as processes that are independent from another yet occur at the same time. In relation to the Uncanny Valley this illustrates strikingly how visual cues can build up along the hierarchical axis and conclude in the detection of a face. Based on the alleged detection of a human face, expectations about social interaction are elicited which then begin to conflict with the accumulating, newly observed and inconsistent cues. Yet, this finding is not restricted to artificial faces but provides ground to regard the Uncanny Valley as a phenomenon of conflict between bottom-up and top-down processing.

1.3.2 Face prototype

Lastly, before we can identify a face, we must detect it. In order to do so, we must first have a face prototype such that we can compare the viewed stimuli to it and judge if it falls into the category face or not. There has been ample evidence of the emergence of face prototypes in young infancy and how they influence our perception throughout our whole life. Le Grand, Mondloch, Maurer, and Brent (2001) conducted a study on children’s proficiency in configural and feature-based processing with young infants with visual impairments and an age-matched control group with normal subjects. They found that the children who were deprived of patterned visual input from birth until 2-6 months of age suffered permanent deficits in configural face processing, however remained perfectly sensitive towards changes of features in faces (Le Grand, Mondloch, Maurer, & Brent, 2001). A possible explanation for this is that young infants must rely on configural visual input given their low visual acuity in the early months of age. In a similar fashion, it was observed that if children are exposed to primate faces from 6 to 9 months of age their discrimination ability for these species remains high, whereas children with no such exposure show a decline in discriminatory ability after 9 months of age (Pascalis et al., 2005). Thus, it can be said that young infancy is a critical period in which we develop a face prototype that predominantly relies on the configuration of faces such as the contour and location of facial features as shown in Le Grand, Mondloch, Maurer, and Brent (2001) and that increases our discriminatory ability for the encountered prototypes (Pascalis et al., 2005). This is again in line with the functional plasticity observed in more specialized brain areas of the visual cortex (Grill-Spector & Malach, 2004). A more subtle instantiation of the same effect is found in what is called the other-race effect or other-species effect (Pascalis et al., 2005). It describes a lower discrimination ability for members of ethnic groups other than the group one grew up with or completely different species (Pascalis et al., 2005). However, if a child is adopted from 3 years of age onwards the effect is not observed and it can similarly be extinguished with intensive training through frequent exposure in adults although this was only proven for the other-race effect. Notably, the study of Pascalis et al. (2005) showed that, at least in infancy, even face prototypes of different species can be learned. This further proves the plasticity of face prototypes and their effect on our perception.

1.3.3 Two paths to the Uncanny Valley: Slow and Fast evaluation systems

Prior to the explanation of the two categories of theories, a subtle distinction should be made. Both approaches offer a different path into the uncanny valley. The slow system provides cognitive models for how the feeling of eeriness comes to be, yet it fails to provide a reason for why the feeling occurs other than it being the implicit by-product of cognitive dissonance. In contrast, the fast system gives elaborate explanations on potential origins of the uncanny valley, by relating it to hypothesized mechanisms that might have improved our reproductive fitness in the past. On the downside however, it lacks explanatory power to describe how the feeling emerges. Yet, the main difference remains the difference in duration the theories of each category require. This is due to the slow system focussing on cognitive processes of higher complexity that include for example learned and explicit categories and therefore require more time. In opposition the basis of the fast system theories is that these processes evolved over the course of evolution. Consequently, the individuals with the quickest reaction time had the greatest advantage for survival and were more likely to pass its genes on to the next generation (reproductive fitness). As a result, the theories of the fast system include mechanisms that occur incredibly fast as a result of millions of years of natural selection.

1.3.3.1 Slow system

Violation of expectation hypothesis

Mori et al. (2012) proposed the violation of expectation hypothesis himself based on his personal experience of eeriness in response to touching a prosthetic hand. He argued that the discrepancy between visually derived information and tactile feedback caused the emotional response (Wang, Lilienfeld, & Rochat, 2015). Building onto this observation, Mori et al. (2012) proclaimed the feeling of eeriness will intensify with greater perceived discrepancy and subsequently proposed that if an entity is moving this will amplify the Uncanny Valley effect. In addition to motion, Brenton, Gillies, Ballin, and Chatting (2005) argued that any mismatch between graphical and behavioural realism such as a mechanistic voice of a human-like android would strengthen the experienced eeriness (Mitchell et al., 2011). Lastly, MacDorman, Green, Ho, and Koch (2009) summarized the common factor of the individual approaches and argued the uncanny response occurs if an entity elicits expectations of human qualities yet fails to live up to those expectations. Regarding the outlined interaction between bottom-up and top-down processing in our visual perception, it seems likely that a mismatch between bottom-up and top-down processing results in an experience of dissonance (see 1.3.1). However, the violation of expectation hypothesis seems to provide an incomplete explanation of the Uncanny Valley. Regarding the increasing number of studies that successfully replicated the phenomenon using images, there is little room within the initial expectations can be violated. Consequently, a more general theory has been proposed as described in the following paragraph.

Categorical uncertainty hypothesis

The categorical uncertainty hypothesis has become one of the most prominent theoretical explanations on the Uncanny Valley (Mathur & Reichling, 2016). In contrast to the mismatch between expected and observed human-qualities it focuses on the inability to determine whether the depicted entity is human or not. Jentsch (1906/1997) first described category confusion, stating that the inability to establish the category membership of an entity produces a negative emotional response of unease. Furthermore, the strength of the emotional reaction was claimed to correspond to the level of confusion about the category identity (Mathur & Reichling, 2020). This confusion is greatest at the category boundary given that the boundary represents the point where stimuli are most deviant from their respective category prototype (Feldman, Griffiths, & Morgan, 2009). The perceptual magnet theory explains how this deviance reduces generalization towards the category prototype and therefore increases sensitivity towards even subtle imperfections (see 1.3.1). Subsequently, based on their deviance to the category prototype and our increased perceptual sensitivity, the category identity of stimuli at the category boundary is most difficult to determine and the boundary relates to the most negative emotional response.

To test this hypothesis Mathur and Reichling (2020) compared if the strongest negative emotional response, which is located at the trough of the Uncanny Valley, overlaps with the perceived category boundary of participants. However, in opposition to the hypothesis, category boundary and trough did not overlap. Consequently, Mathur and Reichling (2020) concluded that the categorical uncertainty hypothesis is not sufficient to explain the Uncanny Valley effect and might just be an epiphenomenon.

Realism inconsistency

Lastly, the realism inconsistency theory offers an explanation detached from the category boundary. Despite being founded on the same premise as the categorical uncertainty hypothesis it makes an important subtle difference. Whereas the categorical uncertainty hypothesis argues that getting closer to the category boundary increases sensitivity, the realism inconsistency theory argues that the increase of realism and human-likeness makes observers less tolerant to imperfections (Mathur & Reichling, 2020). Realism inconsistency can refer to the multi-modal mismatch of visual and behavioural appearance as outlined in the violation of expectation theory, yet also extends towards inconsistent levels of realism of facial features such as the eyes, mouth, and overall skin quality (MacDorman & Chattopadhyay, 2016). The previously described scrutiny towards the observation of human-like faces and social situations (see 1.3.1) supports how even minimal imperfections could quickly cause an experience of dissonance. Similarly, Moore (2012) argued that such conflicting cues could create a perceptual tension which can be observed in the results of fMRI scans. Importantly, the realism inconsistency theory proposes that eeriness can be elicited at any level of human-likeness and independent of the category boundary. However, regarding the increasing sensitivity towards imperfections at higher levels of realism, the realism inconsistency theory postulates that the point of stagnating and plummeting likeability represent a level of realism at which humans become especially perceptive of and increasingly perturbed by inconsistent levels of realism.

1.3.3.2Fast system

Threat avoidance hypotheses

In the initial description of the Uncanny Valley Mori et al. (1970/2012) already speculated the effect to be a mechanism of self-preservation. Explanations that build on this notion can collectively be referred to as the threat avoidance hypothesis and mainly include the following two theories (Zhang et al., 2020). First, the pathogen avoidance theory refers to the Uncanny Valley as an evolved behaviour of disease avoidance (MacDorman & Ishiguro, 2006). Inspired by the theory of disgust from Rozin and Fallon (1987) it argues that the feeling of eeriness is rooted in the basic emotion of disgust and elicited upon detecting imperfections in robotic faces which are perceived as indications of diseases or genetic defects (Wang, Lilienfeld, & Rochat, 2015). Given that the genetic similarity of two organisms increases their likelihood to carry transmittable bacteria and viruses, the pathogen avoidance theory seems to provide a promising explanation to why more human-like entities are met with more caution (MacDorman & Ishiguro, 2006). Consequently, to preserve our reproductive fitness, our ancestors developed a special sensitivity towards human-like entities that deviate in some way from the face prototype established based on our surrounding. This sensitivity then causes us to be cautious by eliciting an intense feeling of aversion while more distant species are evaluated much more tolerantly.

Secondly, the mortality salience theory proposes that robots remind us of our own mortality and thus the Uncanny feeling is caused by our fear of death (Wang, Lilienfeld, & Rochat, 2015). The theory is based on the terror management theory which states that conscious and unconscious thoughts of death elicit defence mechanisms. Consequently, the aversive reaction towards Uncanny robots presents our initial reaction to cope with the immediate fear of death (MacDorman & Ishiguro, 2006). Proponents of the theory provide many examples of how robots remind us of our mortality. First, the sudden shutting down of robots might create the impression of death (Hanson, 2006). Secondly, states of disassembly and the revelation of robots' mechanic interior is postulated to induce the fear that humans too are soulless machines (MacDorman & Ishiguro, 2006). Lastly, they argue that the abrupt movements of humanoid robots cause fear of losing bodily control (Wang, Lilienfeld, & Rochat, 2015).

Evolutionary aesthetics hypothesis

A different evolutionary explanation is presented by the evolutionary aesthetics hypothesis. In contrast to the previous two theories, this approach postulates the underlying mechanism of the Uncanny Valley effect increased the reproductive fitness of our ancestors not by facilitating the avoidance of threats but by aiding in the identification of physical attributes of fertility and health (Hanson, 2006). Respectively, the feeling of eeriness would have caused aversion towards individuals that did not show such attributes and consequently seemed unable to withstand the selection pressure (Wang, Lilienfeld, & Rochat, 2015). Strong support for this hypothesis comes from studies demonstrating the universality of criteria among humans to evaluate attractiveness (Hanson, 2006; Cunningham et al., 2002; Zhang et al., 2020). Generally, skin quality and the affinity towards averageness like bilateral face symmetry are outlined as consistent characteristics of attractiveness (Zhang et al., 2020; Green, MacDorman, Ho, & Vasudevan, 2008). However, Hanson (2006) pointed out that the most attractive faces deviate from the average in facial features associated with sexual maturity. This confirms both the existence of an evolved, universal evaluation of the attractiveness and the presence of selective pressure which relates the distinctiveness of features of sexual maturity to higher attractiveness. Green, MacDorman, Ho, and Vasudevan (2008) provided confirmation for the evolutionary aesthetics hypothesis based on the following two observations. First, attractiveness ratings after

13ms of presentation time significantly correlated with previously assessed levels of attractiveness. Secondly, people can consistently remember the attractiveness of a face yet frequently fail to recall what features determined to their judgement. This highlights the assessment of human attractiveness as a rapid and stimulus-driven process which bypasses conscious evaluation. Lastly and in direct relation to the Uncanny Valley, Hanson (2006) examined the relation between attractiveness and eeriness. In favour of the evolutionary aesthetics hypothesis, he found that faces of high attractiveness consistently rated low in eeriness (Hanson, 2006). Overall, the mentioned research provides evidence on how the Uncanny Valley effect might be a product of the evolved mechanism to constantly evaluate individuals' attractiveness to find mating partners of high reproductive fitness.

1.4 Human-likeness

In their critique of the methodological limitations of current studies on the Uncanny Valley Zhang et al. (2020) also mentioned the absence of a unified definition of human-likeness. Lacking a conceptualization of the predictive variable of the Uncanny Valley means we neither know with certainty what factors make a face convincingly human-like or fall off into the Uncanny Valley. So far studies tried to quantify the human-likeness using morphing, implicit human-likeness ratings, and analyses of morphometrical similarity (Moll & Schmettow, 2015; Mathur & Reichling, 2016; Rádlová, Landová, & Frynta, 2018). While morphing produces a single continuous variable describing the human-robot blend it falls short by creating unnatural distortions and thus potentially distorting the Uncanny Valley effect too (Kätsyri, Förger, Mäkäräinen, & Takala, 2015; Wang, Lilienfeld, & Rochat, 2015). Rating stimuli's human-likeness based on implicit judgements seems to provide a reliable measurement, however it obstructs the exploration of the concept of human-likeness. As already mentioned (see 1.1), Rádlová, Landová, and Frynta (2018) demonstrated the potential explanatory value of creative study designs to understand the dimension of human-likeness. In extension to the study of Rádlová, Landová, and Frynta (2018) and inspired by the pathogen avoidance theory, phylogenetic similarity might provide an alternative conceptualization to human-likeness than phenotypical similarity. Consequently, the present study will employ the ancestral closeness of the species within the experimental setup as a second independent variable. The ancestral closeness of each stimulus will be determined by its phylogenetic similarity to the homo sapiens based on the last common ancestor theory (see 2.2.3).

1.5 Research question

The extensive description of the functioning of the human visual cortex allows us to understand how theories such as the categorical uncertainty hypothesis or violation of hypothesis are not processes particular to synthetic characters such as robots or CG animations. It highlights how our visual system is primed towards the detection of faces and relies on a combination of bottom-up and top-down processing to efficiently detect and recognize faces. Therefore, any mismatch within this interaction would cause a form of cognitive dissonance which could result in the experience of negative emotions. Similarly, the evolutionary theories argue that the effect must have originated long before the invention of any form of realistic human-like automata and with recent evidence of the universality of the Uncanny Valley effect these evolutionary approaches gained tremendous support (Koopman & Schmettow, 2019). Lastly, the possible dimensions of the human-likeness variable still remain a mystery which makes it increasingly difficult to further pinpoint the factors responsible for the perceived eeriness. Consequently, this study will exploratively investigate if the effect is indeed independent of artificial faces and instead rely on biological faces to further support the evolutionary explanation. This resulted in the following two research questions:

1. First, we examine if the uncanny valley effect can be replicated using faces of human and non-human primates.

2. Secondly, we explore if phylogenetic similarity, based on the last common ancestor theory, provides a predictive variable for the perceived eeriness according to the uncanny valley.

2. Methods

2.1 Procedure

The survey was accessible on Qualtrics (see 2.5) either directly through a link shared by the researchers or indirectly by participation via SONA (see 2.6). Participants were first presented with a brief introduction about the content, procedure, and duration of the study. On the next page, participants were provided with informed consent and made aware that their participation is voluntary, and they can resign from the study at any point without stating a reason or suffering negative consequences (Appendix B). Participants had to agree to the informed consent by checking the according multiple-choice box. Afterwards, an overview of stimuli representative of the range of possible faces (see title page) was shown and participants were asked to take up to two minutes to make themselves familiar with the images.

Upon proceeding to the next page, the participants were informed of the concrete procedure of the study. Each image was displayed for two seconds whereafter two analogue rating scales were presented. Participants were not told that the first scale remains the same throughout all stimuli and investigates the likeability of the face whereas the second scale randomly presents one of the five item-pairs of the eeriness index of Ho and MacDorman (2017). Before beginning the study, participants were again reminded that the study consists of 4 blocks containing 25 images each and that after every block an opportunity to take a break is provided. Lastly, the participants were asked to read the scale labels carefully and adjust the language to the language they are most proficient in since a correct understanding of the scale labels is crucial. After each block, a page indicating that the participants successfully completed the prior block and can take a break if needed such that they feel well-rested before continuing into the next block. When the last block was completed three short personality questionnaires had to be filled out after which a debriefing revealed the context and research question of the study. In case participants were interested in the results of the study or wanted to make comments they were provided with the e-mail address of one of the researchers and a text box to make remarks.

2.2 Material

To create the study the web-application software Qualtrics (https://www.qualtrics.com/) was used. It is software specially designed for the development of online studies.

2.2.1 Stimulus collection

A stimulus set of 100 stimuli was created for the study including 89 images of biological faces and 11 images of robot faces from the study of Koopman and Schmettow (2019). The robotic stimuli were included as reference points for human-likeness given their predetermined human-likeness score of previous studies (Koopman & Schmettow, 2019). Prior to the selection of biological faces inclusion and exclusion criteria were defined based on the criteria from Mathur & Reichling (2016) (Appendix C). The majority of stimuli were collected from the catalogue of hominid busts of John Gurche (http://gurche.com/) and the open access databases Global Biodiversity Information Facility (https://www.gbif.org/) and PrimFace (https://visiome.neuroinf.jp/primface/). The remaining stimuli were collected through free stock image websites and targeted google image searches to include the faces of different human ethnicities and different displayed emotions. Altogether 111 biological faces were collected. Human-likeness of the faces was rated individually by all four researchers and analysed using intraclass correlation coefficient. The images with the lowest inter-rater agreement were removed from the set to achieve the final set of 89 biological stimuli.

2.2.2 Stimulus preparation

To minimize confounding influences all images were cropped to show only the face and put on white background using Adobe Photoshop. The cropped images showed from the chin upwards and with facial hair, if applicable. Furthermore, all faces were adjusted in size and centralized. Lastly, all images were exported with a resolution of 450x450 pixel.

2.2.3 Stimulus grouping

All stimuli were divided into 10 groups based on how closely related they are to the homo sapiens to compute the variable of Ancestral Closeness. The stimuli were categorized using the last common ancestor theory (Most recent common ancestor, 2021). Thus, each primate was allocated to a group corresponding to its last common relative to the homo sapiens as outlined by contemporary taxonomies (Primate, 2021). A description of the groups and all included species are provided in Appendix D. According to the results of Rádlová, Landová, and Frynta (2018) the ten groups were additionally summarized in four categories of morphometrical similarity. Category 2-4 were directly copied from the results of Rádlová, Landová, and Frynta (2018) whereas category 1 was added to the grouping of the original article. Subsequently, the stimuli set of the current study included stimuli of higher human-likeness and cut off the lower end of the human-likeness spectrum in comparison to Rádlová, Landová, and Frynta (2018). An overview of the number of stimuli for each group can be seen in Table 1. It should be noted that the group numbers of the Ancestral Closeness are coded reversely to human-likeness with the most closely related human ancestors receiving lower numbers and the distant human relatives having higher numbers.

Group Ancestral Closeness	0	1	2	3	4	5	6	7	8	9	10
Category	-	1	1	1	2	2	2	2	3	4	-
Number of stimuli	10	3	11	16	16	9	4	11	5	2	11

Table 1. Frequency of stimuli for each category

2.3 Participants

In total 82 participants took part in the study. All participants were recruited through convenience sampling from either the social environment of the researchers or from the test subject pool SONA of the Behavioural, Management and Social Science faculty at the University of Twente. Participants from the test subject pool were rewarded with credits required for the completion of their studies. The study was conducted on the premise that the Uncanny Valley effect is a universal experience, thus no demographical data was collected (Keeris & Schmettow, 2016). Requirements for participation in the study were being at least 18 years old, proficient in English and without any major visual impairment. Given that the analysis does not suffer from missing values no participants had to be excluded for this reason (Schmettow, 2021). However, four participants provided too little data points to compute an individual polynomial.

2.4 Measures

To assess participants' reactions toward the stimuli two visual analogue scales were implemented in the survey. The instructions for these scales read "To me this face seems…" followed by the respective scale. First the likeability scale from Mathur & Reichling (2016) was presented which ranges from '-100' (less friendly, more unpleasant, creepy) to '+100' (more friendly and pleasant, less creepy). The second scale randomly presented one of the five item-pairs from the spine-tingling subscale of the eeriness index of Ho and MacDorman (2017). Participants could respond on a scale ranged from '0' to '100' and the respective item-pairs are 'Uninspiring – Spine-tingling', 'Boring – Shocking', 'Predictable – Thrilling', 'Bland – Uncanny', and 'Unemotional – Hair-raising'. The eeriness index was selected as a measurement given its good psychometric properties, showing a high internal reliability with an Cronsbach's alpha of .84, and successful application in previous studies (Ho & MacDorman, 2017; Haeske & Schmettow, 2016). Lastly, all instructions and items related to the measures were provided in German and Dutch in addition to English, to avoid any misunderstanding due to language difficulties (Appendix E).

2.5 Design

We used a within-subjects design to investigate how the successive eeriness ratings of participants behaved over the range of human-likeness depicted by our stimuli set. Consequently, the responses of participants on the eeriness scale were used as the dependent variable whereas the predetermined human-likeness of each stimulus was included as the independent variable in the model. Ancestral closeness and emotional valence are additional independent variables serving as an alternative for human-likeness and a control variable respectively. For more information on the variables see section 2.2.3 and 2.4.

2.6 Data analysis

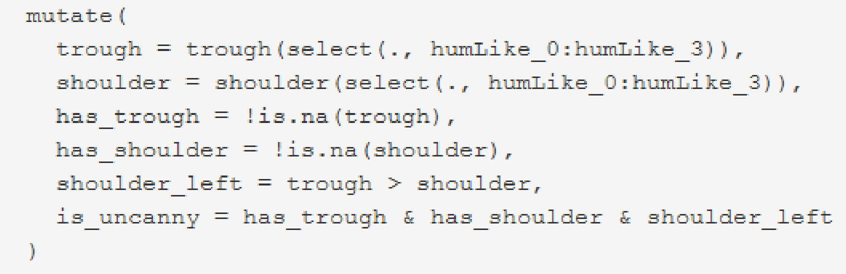
For the data analysis, the raw dataset had to be slightly transformed. First, participant's responses on the eeriness scale were rescaled by a factor of .999. This allowed to analyse the continuous data from the visual analogue scales using a beta regression which requires responses to be strictly within an interval from 0 to 1. Additionally, the responses on the scale were multiplied by -1 such that a high score indicates high likeability and low eeriness and vice versa. The individual human-likeness ratings of each researcher for the stimuli showed very good interrater reliability (>.92) (Appendix F). Thus, the mean human-likeness was computed and included into the dataset. Interestingly, ancestral closeness was also highly correlated to the individual human-likeness ratings (>.77) and thus not included as a separate variable in the analysis (Appendix F). All exact correlations can be found in Table 2. Furthermore, emotional valence was included as a control variable. Lastly, two filter were applied such that only the responses on the eeriness scale and the faces of non-human primates were included in the main analysis.

2.6.1 Analysis on population-level

For visualization purposes, a simple regression analysis was computed using the averaged human-likeness ratings as the independent variable and the averaged eeriness ratings of all participants for each stimulus as the dependent variable. For a more accurate insight on the relation between human-likeness and eeriness on population-level, four polynomial models were computed using the Markov chains Monte Carlo (MCMC) method (Schmettow, 2021). The polynomial functions were of degree 0 (grand mean), degree 1 (linear), degree 2 (quadratic), and degree 3 (cubic). All polynomial models further included emotional valence as a control variable.

To evaluate which of the four models represents the data best, the relative predictive accuracy for each model was estimated using the LOO-IC (Schmettow, 2021). Usually, cross-validation is regarded to be the golden standard for this, however, it comes at the cost of separating our data into two parts. The first one is used to train the model whereas the second one must be left out of the model estimation such that we have sample data independent of the model to evaluate its predictive accuracy. Fortunately, the leave-one-out (LOO) method circumvents this issue by iteratively excluding only one observation as sample data and allowing us to train the model with minimal compromise to the dataset (Schmettow, 2021). In comparison to the LOO, the LOO-IC provides a more computation efficient analysis. Nonetheless, it accounts for the model fit, referring to the discrepancy between the idealized (predicted) response of the model and the observed response in the data, and furthermore controls for over-fitting by taking the complexity (quantified as the number of parameters) into account (Schmettow, 2021). The output of the LOO-IC is a single value, the information criteria (IC). It is derived from the addition of deviance and complexity of the respective model and thus a lower IC indicates an overall better predictive accuracy in relation to model complexity (Schmettow, 2021).

In addition to the model comparison using LOO-IC, we can compute the probability that the Uncanny Valley is present within the data in a more direct fashion. For this we take the characteristics of the Valley curvature as condition for its existence and check for the conditions in each iteration of the MCMC of the third-degree polynomial (Schmettow, 2021). Consequently, regarding the Valley curvature, we must examine if two stationary points are present, which of these points is a local maximum and which is a local minimum and lastly if the maximum (the shoulder) precedes the minimum (the trough). To identify stationary points of the polynomial and determine whether they are minima or maxima we use the first and second derivative respectively. The whole procedure can be seen in Fig. 2 (Schmettow, 2021). The probability of the Uncanny Valley is computed based on the percentage of MCMC samples which fulfil the conditions in relation to how many samples do not.



```
mutate(
  trough = trough(select(., humLike_0:humLike_3)),
  shoulder = shoulder(select(., humLike_0:humLike_3)),
  has_trough = !is.na(trough),
  has_shoulder = !is.na(shoulder),
  shoulder_left = trough > shoulder,
  is_uncanny = has_trough & has_shoulder & shoulder_left
)
```

Fig. 2 Conditions for confirming the Uncanny Valley curvature

2.6.2 Analysis on participant-level using Multilevel modelling

The application of Multilevel models allows us to regard the data on participant-level in direct comparison to the population-level (Schmettow, 2021). The most profound implication of this is that each participant gets its own coefficient and thus introduces an individual factor which enables us to observe the individual variance in response styles throughout participants and to investigate the universality of the Uncanny Valley (Schmettow, 2021).

Again, we compute a model to describe the data using the MCMC method and visualize it on a graph to get an overview of the data. However, instead of taking the average eeriness ratings, we group by participants and thus compute an individual polynomial for each participant (Schmettow, 2021). Next, we use the Bayesian regression model using STAN (brms) which allows to include multiple formulas in our regression model (Schmettow, 2021). Through this, we can apply a distributional analysis which relaxes the assumption of equal variance of the generalized linear model and simultaneously includes response variance as a variable in the analysis. Response variance can be a serious threat to the analysis of rating scales. This is generally referred to as the issue of anchoring and describes how participants utilize rating scales and their range differently based on their subjective interpretation of the extremes (lower and upper endpoints of the scale) (Schmettow, 2021). Thus, by accounting for response patterns we achieve an improved approximation (shown by an increased IC) of the cubic polynomial on participant-level (Schmettow, 2021). To double check and evaluate the model fit we compared the predictive accuracy for the model with and without the distributional analysis based on the LOO-IC.

Lastly, we can compute the probability of the Valley curvature again using the above-mentioned conditions with the only difference that we do the computation on participant-level. The results are reported in two ways. First, the probability of the existence of the Uncanny Valley is reported for each participant. Secondly, the individual positions of the trough and shoulder which result as a by-product of the first computation will be reported to further visualize the universality or individual differences between participants (Schmettow, 2021).

3. Results

In this study, we examined if the Uncanny Valley can be replicated using only primate faces. This was tested twice. First, on population-level by evaluating which polynomial, of 0th degree up to 3rd degree, presented the best predictive accuracy based on a model comparison analysis using IC and by computing the probability of an Uncanny Valley curvature using the characteristics of the curvature as conditions. Secondly, a similar analysis was run on the participant-level through the use of multilevel modelling to analyse the probability of the Uncanny Valley curvature for each participant separately. Furthermore, the ancestral closeness of stimuli was determined as an alternative independent variable to human-likeness (see 2.2.3). However, a correlation analysis between the individual human-likeness ratings and ancestral closeness revealed them to be congruent (Table 2). Consequently, ancestral closeness was not included in the following analyses but can be assumed to produce the same results as the analysis using human-likeness.

	Rater 1	Rater 2	Rater 3	Rater 4	AncestralCloseness
Rater 1	NA				
Rater 2	0.9554019	NA			
Rater 3	0.9436330	0.9303407	NA		
Rater 4	0.9328772	0.9426051	0.9438880	NA	
Ancestral Closeness	-0.7707424	-0.9422645	-0.9379207	-0.9268578	NA

Table 2. Correlations of human-likeness ratings between 4 raters and Ancestral Closeness

3.1 Population-level

The regression plot on population-level indicates a curvature similar to the Uncanny Valley. However, it lacks a clearly distinguishable shoulder and the averaged responses appear quite scattered (Fig. 3). Concurringly, with regard to model complexity, the cubic model is not preferred to represent the data based on the LOO-IC. Out of the four polynomial models, the linear model showed the best predictive accuracy on population-level (Table 3). Consequently, the probability of the Uncanny Valley curvature on population-level turned out moderate (.6257).

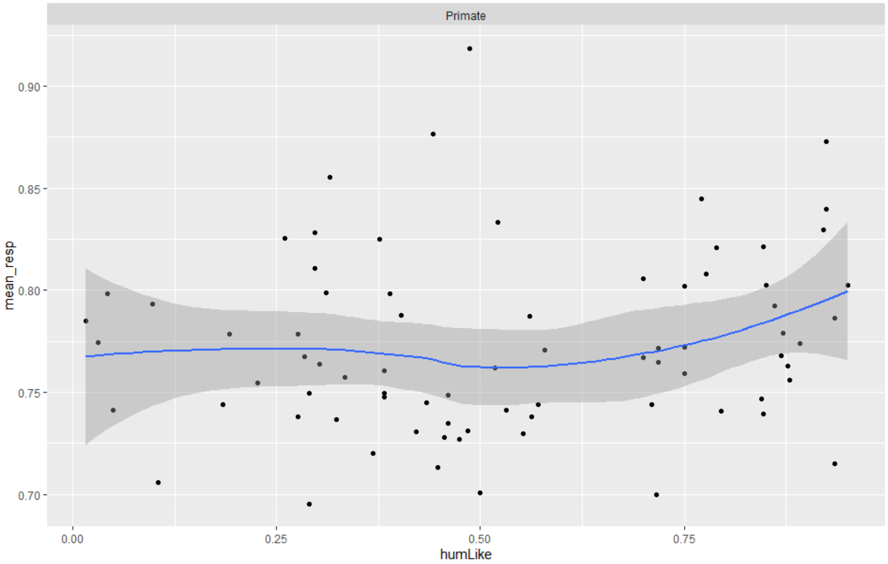


Fig. 3 Regression plot between human-likeness (x-axis) and averaged responses on the eeriness scale (y-axis)

Model	IC	Estimate	SE	diff_IC
Linear model (1st degree)	looic	-268.54	12.32	0.0000000
Grand mean (0th degree)	looic	-268.05	12.12	0.4824929
Quadratic model (2nd degree)	looic	-266.25	13.16	2.2800595
Cubic model (3rd degree)	looic	-264.38	13.17	4.1594711

Table 3 Polynomials ranked by predictive accuracy based on LOO-IC

3.2 Participant-level

With the Multilevel model, we can examine each participants' individual responses in relation to the population. As visible in Fig. 4, the Uncanny curvature is much more distinct for individual responses. Appendix G entails separate plots for each participant for illustration. Furthermore, the multilevel model illustrates that the results on population-level are inconclusive. The range of intercepts between the individual polynomials of participants indicates anchoring differences within the sample population. Consequently, the averaged data on population-level is distorted through the variance in response patterns of participants.

To account for the differences in response style and thus achieve a more accurate model we run the multilevel model again with the distributional analysis included Fig. 5. The new model showed a much better predictive accuracy as indicated by the lower IC shown in Table 4.

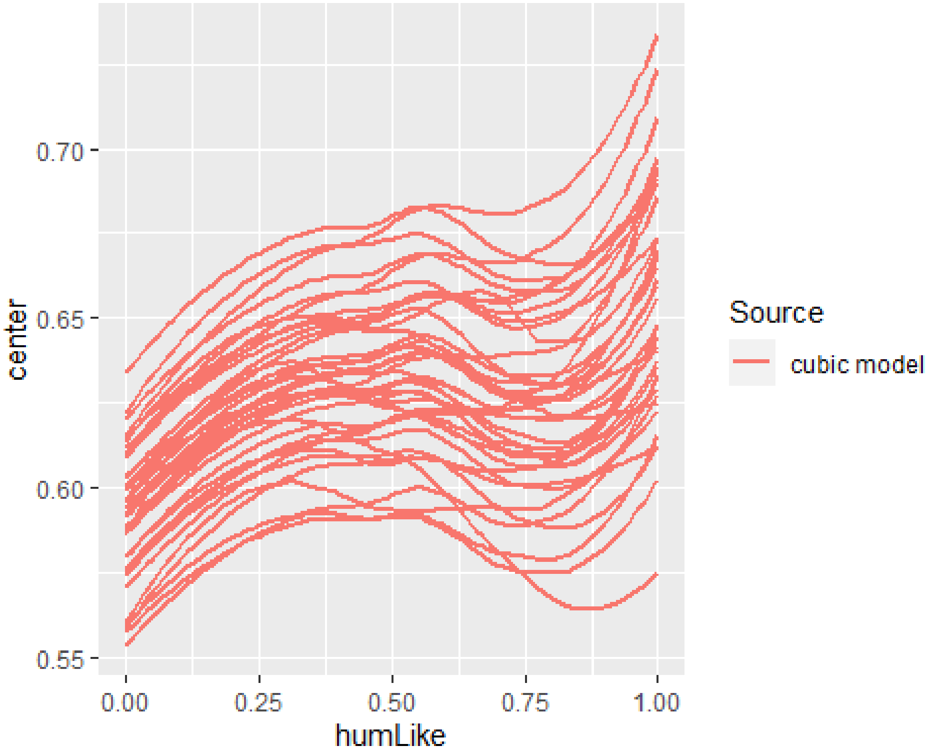


Fig. 4 Multilevel model showing the range of individual responses by separate polynomials for each participant

Model	IC	Estimate	SE	diff_IC
Multilevel model with distributional analysis	looic	-5632.258	153.9320	0.000
Multilevel model without distributional analysis	looic	-2226.516	159.7946	3405.742

Table 4 Comparison of predictive accuracy for multilevel model with and without distributional analysis based on LOO-IC

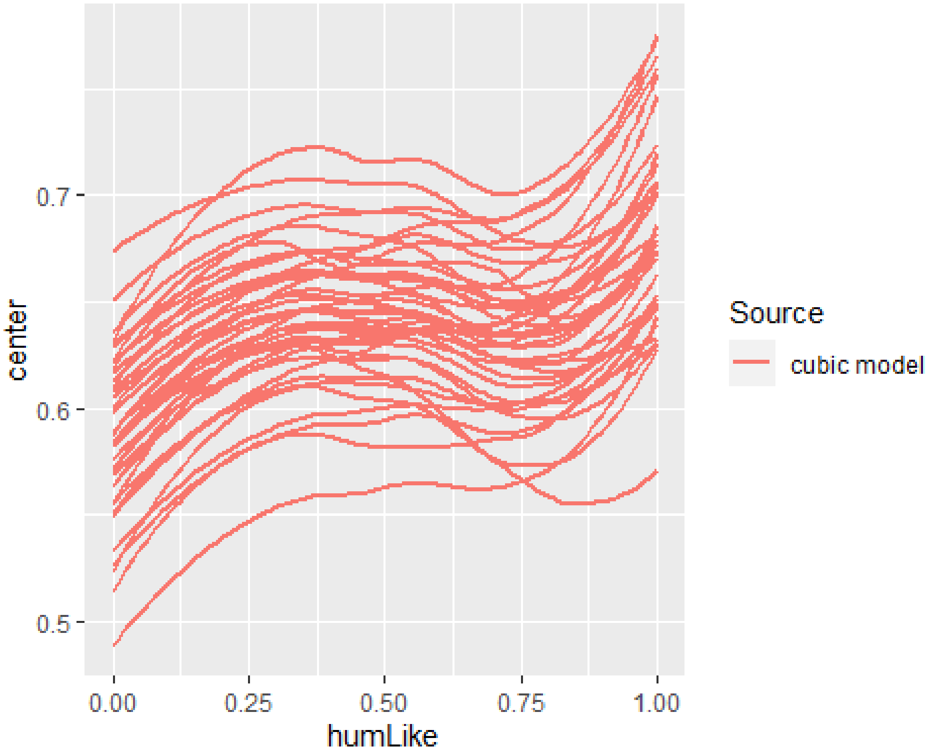


Fig. 5 Multilevel model with distributional analysis showing the range of individual responses by separate polynomials for each participant

Lastly, by application of the multilevel model we investigate the universality of the Uncanny phenomenon within the dataset. Almost every participant showed a high probability to fall into the Uncanny Valley (Fig. 6). Only 7.7% of participants showed a probability <.8 to exert the Uncanny effect and 84.6% of participants experienced the phenomenon with a certainty >.9. Furthermore, the individual positions of shoulder and trough per participant revealed extremely high concurrence for the position of the shoulder (Fig. 7). This is a very striking observation for the universality of the effect as it shows that not only does the effect occur consistently throughout participants, but moreover it is elicited at the same degree of human-likeness for individual observers.

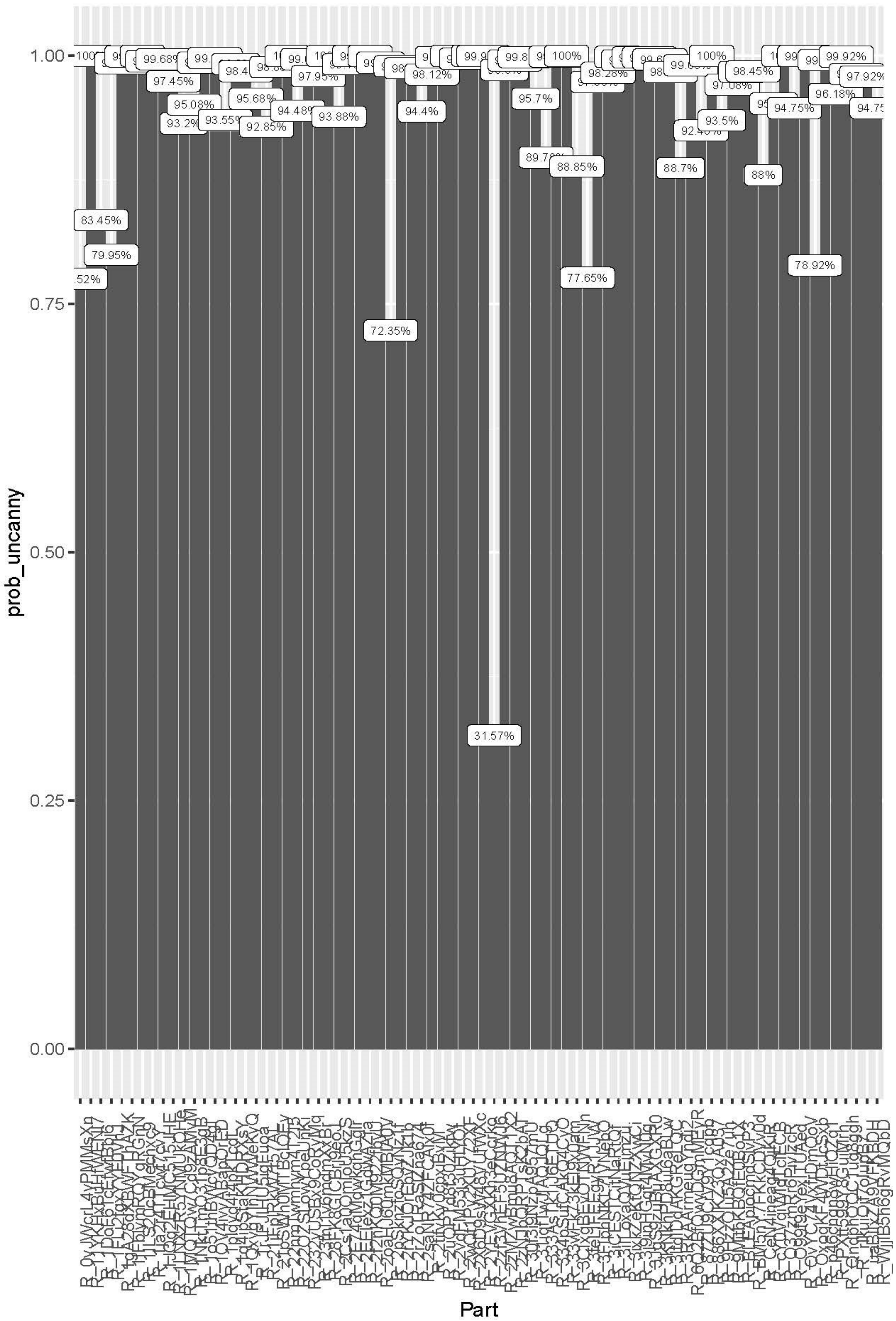


Fig. 6 Probability to exert the Uncanny Valley curvature for each participant

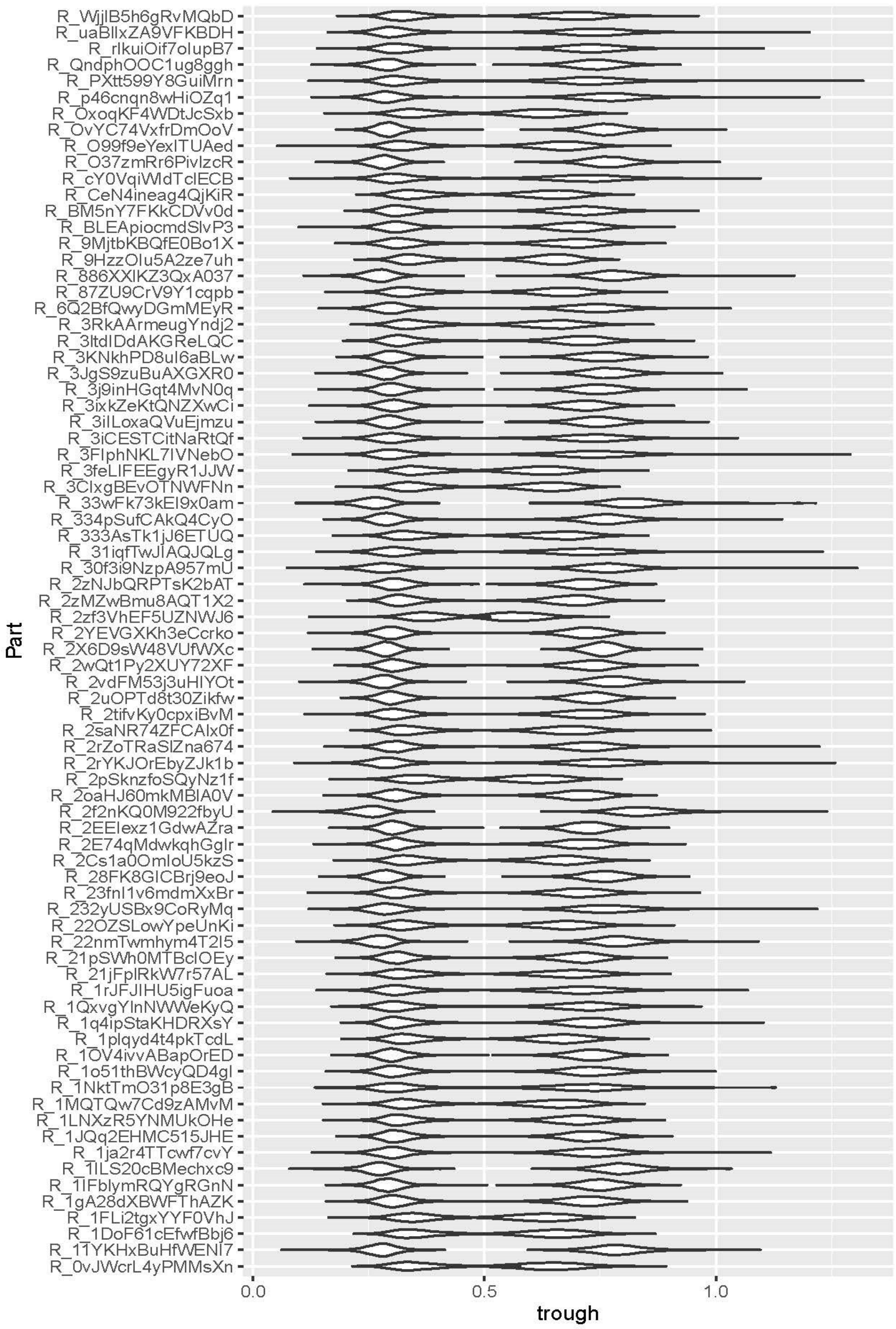


Fig. 7 Position of shoulder and trough are shown separately for each participant

As already indicated few participants did not show a high probability towards exerting the Valley curvature. Especially one participant presents an outlier in the data, showing rather a linear positive correlation with human-likeness and the absence of a valley (Fig. 8). This naturally speaks against the universality of the effect. Moreover, it must be noted that the computation of the Uncanny Valley probability is prone to overestimate the likelihood of the effect. As it is based merely on the condition that a local maximum must be followed by a local minimum it confirms even slight extreme points as Uncanny Valley curves. This includes polynomials that lack a distinct shoulder or trough and thus neither indicate an abrupt negative affective response (eeriness) nor that unrealistic faces are perceived as more likeable than highly realistic ones as presumed in the hypothesis of the Uncanny Valley (Fig. 9). Lastly, some participants showed curves in which the low human-like stimuli were rated as eerier than stimuli within the valley. This contradicts the original conception of Mori et al. (1970/2012) and the 'Uncanny' Valley which hypothesized that it is the high level of human resemblance that causes an especially negative reaction. In contrast curves like the one depicted in Fig. 10 argue for a more negative response to low human-like stimuli.

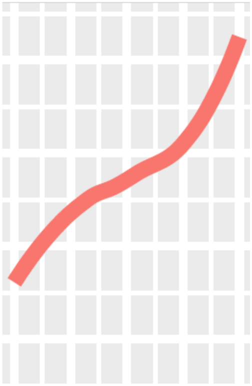


Fig. 8 Participant without Uncanny Valley curvature

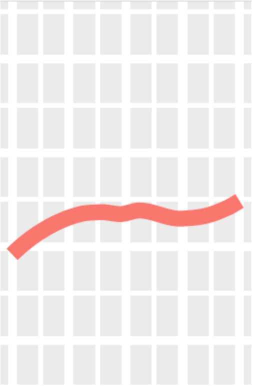


Fig. 9 Lack of distinct shoulder and trough

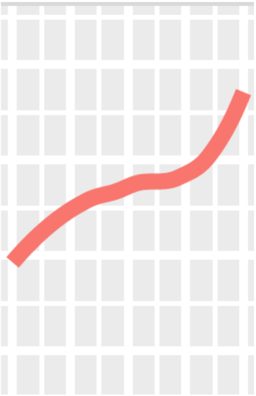


Fig. 10 Low eeriness within trough but relatively high eeriness for lower human-likeness

4. Discussion

4.1 Interpretation of results in relation to the research question

Despite the vast amount of recent research on the Uncanny Valley and the abundance of theoretical frameworks to explain the effect, it is still unclear what factors underlie the phenomenon. In an effort to narrow down the multitude of possible explanations this study exploratively probed the perseverance of the presented evolutionary notions. The evolutionary theories form one out of two explanatory groups on the Uncanny Valley describing fast, highly specialized and automatic processing in contrast to the hypothesis involving more complex and slower processing that partially also include higher degrees of conscious thought. Given that evolutionary mechanisms evolve slowly over the course of millions of years the evolutionary hypotheses presuppose that the origin of the Uncanny Valley greatly precedes the development of robots or any form of automata for that matter. Therefore, to test the validity of this hypothesis, this study pursued two research questions. The first question investigated if the Uncanny Valley is observable along a different dimension of human-likeness than from artificial or mechanistic faces to humanoid and realistic ones. Thus, we tested if the Uncanny Valley can be replicated with a set of stimuli including only biological faces of human and non-human primates. Furthermore, the second research question examined more specifically how the dimension of human-likeness might be conceptualized other than through implicit ratings of perceived human-likeness. Consequently, we tested if the ancestral closeness, referring to how closely a species is related to the modern human (homo sapiens) based on the last common ancestor theory, presents a predictive variable for perceived eeriness according to the Uncanny Valley.

In this study, we could not replicate the Uncanny Valley on population-level. However, on participant-level, the Uncanny Valley was observed consistently. Furthermore, regarding the participants' individual responses alongside one another allows reinterpreting the results on population-level. First, based on the observation that the Uncanny Valley can be found in almost every participant we can safely conclude that the Uncanny Valley is indeed elicited by primate faces. Secondly, the multilevel model reveals how the inconsistency of results between population and participant-level is due to the different utilization of the rating scales by participants. This primarily shows that the Uncanny Valley effect is not limited to the domain of robotics and engineering in which it was initially discovered by Mori et al. (1970/2012) but that it exists independently from it. Subsequently, this illustrates that rather than being an engineering problem the Uncanny Valley is in fact a more general psychological phenomenon. In regards to the motivation of our research question, our results confirm that it is highly likely that the mechanisms underlying the Uncanny Valley effect arose in a different context and manifested themselves a long time ago as a behaviour to increase the reproductive fitness of our ancestors. Here the results of the second research question provide a lot of room for further speculation. Ancestral closeness correlated highly with the ratings of human-likeness which implies that both variables are congruent and will produce similar results. Consequently, the phylogenetic similarity can successfully be used to replicate the Uncanny Valley curvature. This opens up the debate to suspect the origin of the Uncanny Valley as a particular instance of a mechanism of discriminating conspecifics from different species or to facilitate successful mate selection as we will be discussed in the following.

4.2 A critical look at the present study

Zhang et al. (2020) pointed out a variety of common problems in the methodology of studies on the Uncanny Valley of which some also apply to the present study. First, they mention the lack of a clear definition of the uncanny feeling. In this study, a subscale of the eeriness index from Ho and MacDorman (2017) was used. This index showed strong internal reliability and factor analyses confirmed the index to be uncorrelated with concepts such as attractiveness or humanness. However, in the current as well as in past studies using the eeriness index participants mentioned troubles to interpret the items as they either did not represent the emotion they felt in response to the stimuli or because the item-pairs did not appear as opposites to another (Haeske & Schmettow, 2016; Koopman & Schmettow, 2019). The significance of accurately conceptualizing the uncanny and developing corresponding and understandable indexes for experiments is illustrated by the impact of different labels on participants' responses on rating scales. Different labels can cause differences in the anchoring of participants with more extreme labels causing more conservative utilization of the range of a visual analogue scale while more mild labels encourage respondents to be more daring (Schmettow, 2021). Lastly, a lack of objective indexes impairs the generalizability of experiments (Zhang et al., 2020). In contrast, the observation of the valley curvature in the data suggests that the particular wording of labels does not severely impair the ability of participants to judge a stimulus. Ho and MacDorman (2016) postulated that as long as the scale is confined by a neutral label at one end and a negative one at the other ratings would remain valid. However, it must be noted that for more informative results a variance in the absoluteness and explicitness of labels is desirable.

Mori et al. (1970/2012) originally suggested building less realistic robots as he argued in his initial model of the Uncanny Valley effect that less human-like robots are perceived more favourably than those that fall into the valley. However, when carefully examining the individual polynomials of participants (see Appendix G) it can be observed that oftentimes the low human-like stimuli produce a similar eeriness score as those who fall into the valley. Participants who show these curves either present a shoulder so flat or trough so shallow that the data barely meets the original conception of the Uncanny Valley (Fig. 9). These cases might be explained due to participants anchoring and a very narrow exploitation of the range of the rating scale. However, in some cases, even the opposite of the proposition of Mori et al. (1970/2012) can be seen (Fig. 10) and low human-like stimuli receive higher eeriness ratings than those in the valley. This directly conflicts with the notion of an 'Uncanny' Valley as it shows that faces within the valley do not induce an eerie sensation or at least do not elicit a more negative feeling than the less human-like stimuli. In this regard, many have criticized the Uncanny Valley model. Brenton, Gillies, Ballin and Chatting (2005) for once pointed out that the existence of an uncanny reaction does not itself confirm the valley model and consequently advocated a careful and unbiased interpretation of data using multidimensional scaling. Upon careful investigation, it might be suggested that the data shows an aesthetical plateau, where affinity more or less stagnates at a certain degree of human-likeness before it increases again at higher levels of human-likeness. This eventually results in the issue that there is a lack of conceptualization and consensus on what defines the uncanny response as criticized by Zhang et al. (2020). Revising the current stimuli set to include a higher density of stimuli at the human-likeness positions of shoulder and trough might produce more distinct curvatures that correspond more closely to the model of Mori et al. (1970/2012). An imperfect distribution of stimuli across the dimension of human-likeness was to be expected given that for the purpose of this study a novel stimuli set was created. Additionally, it can be argued that using a repeated measure design and exposing participants repeatedly to the stimuli would result in a more detailed Uncanny Valley curvature as Koopman and Schmettow (2019) were able to observe the Uncanny Valley effect universally throughout all participants using this method. Another approach to investigate this problem is mentioned in 4.5. In contrast, it could also be speculated that the Uncanny Valley effect observed in response to biological faces does not occur as distinct as it does for robotic faces.

Lastly, in face of the scope of this thesis, it is impossible to include all the different theories and explanatory approaches to the Uncanny Valley. Especially many of the slow system theories and explanations based on individual differences were mentioned only marginally. While the results in this study strengthen the hypothesis of an evolutionary explanation to the Uncanny Valley, many slow system theories nonetheless provide a lot of ideas that further the understanding. Consideration of these theories such as linking the Uncanny Valley to phobias like coulrophobia or pediophobia (fear of clowns and dolls respectively) can help to discern potential cofounding influences on the Uncanny Valley and thus create a more holistic picture (Wang, Lilienfeld, & Rochat, 2015). Whereas such phobias are rare it could be postulated that all humans have some degree of fear towards clowns and dolls, after all they too represent human-like artifacts. The phobias would therefore only present extreme cases of these fears. Consequently, it could be fruitful to link the Uncanny Valley to other human-like artifacts or look for inspiration within phobias to investigate the effect more accurately. Lastly, some potentially confounding variables were not controlled for such as the influence of attractiveness (Hanson et al., 2005; Rádlová, Landová, & Frynta, 2018). Rádlová, Landová, and Frynta (2018) critically mentioned that the attractiveness of individual primates might work against the Uncanny Valley effect in response to the observation of outliers in their study examining the attractiveness of different human-like primates. Hanson et al. (2005) likewise argued for the influence of aesthetics, proclaiming that any robot regardless of its human-likeness can be appealing or uncanny depending on its aesthetics and social responsiveness. However, the conceptualization of attractiveness can be a complicated endeavour. Rádlová, Landová, and Frynta (2018) pointed out how species more closely related to the homo sapiens were rated based on different factors as more distant human relatives. Gothard, Brooks, and Peterson (2009) also showed that macaque monkeys use different perceptual strategies for the identification of conspecifics than for discriminating other species. This goes to show that the criteria outlined to determine the evaluation of human attractiveness as for example by Cunningham (1986) or Seyama and Nagayama (2007) cannot simply be generalized towards the attractiveness ratings of primates or other species. Nonetheless, revealing these factors would undoubtly forward the understanding of Uncanny Valley by account for single deviating stimuli or by providing a clearer grasp on what makes faces (dis)likeable.

4.3 Relation to existing research

In the following paragraphs, we will discuss the implications of our results to the particular theoretical explanations on the Uncanny Valley and comparable studies mentioned in the beginning.

Pathogen avoidance hypothesis

The pathogen avoidance hypothesis postulates that the mechanism underlying the Uncanny Valley effect developed over the course of human evolution to cause us to be cautious towards closely related species as they might carry transmittable diseases and gene defects. Therefore, the theory implies that the effect should also be observable in response to biological faces that resemble us humans. Consequently, our results strengthen this hypothesis. Most significantly in favour of the hypothesis is the strong correlation of ancestral closeness with human-likeness. This indicates that not only the implicit ratings of human-likeness but furthermore the phylogenetic similarity of species to the homo sapiens predicts the Uncanny Valley. Hereby we directly show that, as predicted by the pathogen avoidance theory, modern humans still consistently show an aversion towards former relatives of our ancestors. As a conclusion of our results the pathogen avoidance hypothesis presents a very promising explanation to the Uncanny Valley and requires more specific examination.

Evolutionary aesthetics

Like the pathogen avoidance hypothesis, the evolutionary aesthetics hypothesis argues that the Uncanny Valley effect originated as a behaviour to increase our reproductive fitness. However, this theory proposes that an intricate automatic process to identify individuals of high reproductive fitness for mating underlies the effect. Therefore, the aversive reaction does not serve to avoid potentially sick individuals but those unfit for reproduction. Similarly, our results confirm the theory by replicating the effect with biological faces.

Realism inconsistency

After finding contradicting results to the category confusion theory in their recent study, Mathur and Reichling (2020) suggested realism inconsistency as a promising alternative to the category confusion theory. Whereas inconsistent levels of realism might affect the eeriness of a synthetic character, our results strongly oppose the notion that they cause the Uncanny Valley. Given that our study only used unmanipulated biological faces we successfully replicated the Uncanny Valley effect based on consistently real faces.

Mortality salience

Similar to the realism inconsistency theory, the mortality salience theory proposed that synthetic characters remind us of our own mortality due to their ability to shut down, be deconstructed or because of abrupt and inhuman movements (Hanson, 2006; MacDorman & Ishiguro, 2006; Wang, Lilienfeld, & Rochat, 2015). As before, through the use of biological faces, these assumptions fall flat. The theory might nonetheless provide a cofounding variable to the perceived eeriness towards robots, yet it can be concluded that it is not the main cause of the phenomenon.

Replicating the Uncanny Valley with biological faces

The first experiment relating the Uncanny Valley to biological faces came from Rádlová, Landová, and Frynta (2018). With the primate species included in their study which mainly investigated the factors underlying the human evaluation of the attractiveness of primate faces, they were unable to replicate the Uncanny Valley (Appendix F). Based on their results Rádlová, Landová, and Frynta (2018) hypothesized the Uncanny Valley might be observable upon closer investigation of the phylogenetically closest group of primates, the catarrhine. Our results confirm this hypothesis by demonstrating that including more stimuli from the category of the catarrhine and further expanding our stimuli set to include even closer relatives to the homo sapiens (see 2.2.3) showed the Uncanny Valley curvature. This goes not only to show the replicability of the Uncanny Valley with biological faces but furthermore that the stimuli set created in this study can be used to produce the Uncanny Valley effect.

Individual differences in the experience of the Uncanny Valley

MacDorman and Entezari (2015) found a correlation between individual attitudes such as religious fundamentalism, neuroticism, and animal reminder sensitivity and the occurrence of the Uncanny Valley. However, follow-up studies were not able to replicate the results of MacDorman and Entezari (2016) (Haeske & Schmettow, 2016). While this study did not examine participants' attitudes, the observation that the Uncanny Valley effect is elicited consistently at the same level of human-likeness serves as evidence to refute individual difference as the cause for the Uncanny Valley and further that it affects the susceptibility of people for the effect. It can be speculated that the correlations MacDorman and Entezari (2016) observed are due to systematic differences in peoples response styles and utilization of response scales (see 2.6.1) which are correlated with personality differences. What remains unclear from our results is if the strength of the emotional response people experience towards uncanny faces differs with individual differences. This would be represented in the depth of the trough of participants. However, due to the same reason that people differ systematically both in their mean responses and the absolute range of the scale they utilize, an analysis using continuous or even categorical scales (like a 5-point Likert-scale) does not allow to simply compare the absolute nor relative depth of the trough in comparison to the shoulder. As a solution might be presented by qualitative assessment of the participants' emotion (see 4.5).

4.4 Conclusion

Based on our results we fully established a bridge of the Uncanny Valley phenomenon to being a psychological phenomenon and particular manifestation of a more general mechanism. Previous research showing the universality of the effect on participant-level already strongly pointed in this direction however the measurement using artificial faces still related it to the domain of engineering and design. In contrast, replicating the effect with purely biological and unmanipulated faces detaches the effect from the context of robotics and artificial faces allowing us to safely conclude that the effect has presumably much deeper roots within a different setting than the interaction with androids. This allows us to focus more specifically on existing evolutionary explanations of the Uncanny Valley in the future.

4.5 Future research

The current study further supported the notion of an evolutionary origin of the Uncanny Valley as proposed in previous studies (Koopman & Schmettow, 2019; Keeris & Schmettow, 2016; Slijkhuis & Schmettow, 2017; MacDorman, Green, Ho, & Koch, 2009). Most prominent seem the pathogen avoidance hypothesis and evolutionary aesthetics hypothesis. Therefore, follow-up studies should apply exploratory experimental setups to narrow down the context within the mechanisms behind the Uncanny Valley originally emerged. Furthermore, the controversy around the ‘uncanny’ of the Uncanny Valley remains (see 4.2) and it is unclear whether the drop in affinity observed in peoples’ responses corresponds to a feeling of eeriness and strangeness.

As outlined in 4.2 the evolutionary aesthetics hypothesis suffers from a lack of understanding of how the evaluation of attractiveness is determined, especially in response to different species. Further in line with this is the unclear conceptualization of human-likeness which presents no answer to what specific factor make a face more human-like and likeable and which cause the opposite. As a solution, an exploratory study using eye-tracking software could be conducted to compare if the most salient facial features differ throughout conspecifics and other primates. Cheetham, et al. (2013) already conducted such an analysis for the exploration on the Uncanny Valley, yet only compared synthetic and human faces.

A qualitative assessment might reveal more accurate answers to investigate whether participants actually experience an intense feeling of strangeness when confronted with uncanny faces. Such an experiment could utilize a stimuli-set that has been confirmed to replicate the Uncanny Valley effect and includes the positions of trough and shoulder. Participants could then be presented with open questions about the emotion they are experiencing in response to faces within the valley. These open questions could on the one hand be compared with the same question about stimuli located at the shoulder or the lower end of the human-likeness spectrum. On the other hand, it could again be utilized to examine potential differences in the emotional response towards biological faces and artificial faces lying in the Uncanny Valley. Additionally, an analysis of this qualitative data could shed more light on the distortion through systematic differences due to anchoring.

Finally, to investigate whether the mortality salience hypothesis presents a relevant predictor for the Uncanny Valley effect at least in response to robot faces, the experimental setup of MacDorman and Ishiguro (2006) can be replicated. The mortality salience hypothesis was inspired by the terror management theory which claims that reminders of mortality produce a range of attitude changes. More precisely, it states that even if a person is only implicitly and subliminally reminded of their mortality, the elicited emotion is conscious and can be perceived. Therefore, after experiencing a feeling of eeriness in response to a stimulus that lies within the Uncanny Valley people might reconsider their conscious attitudes. Consequently, future studies could use a mixed design measuring changes in the attitudes of participants before and after the presentation of uncanny stimuli. The results can then be compared to a control group which are exposed to regular faces. This study design would shed more light on potential differences of the Uncanny Valley effect for biological and artificial faces.

References

Agora Tec (2018). *The strange robots of uncanny valley*. https://agoratec.com/2018/02/27/strange-robots-uncanny-valley/ Retrieved 15.06.2021.

Brenton, H., Gillies, M., Ballin, D., & Chatting, D. (2005). The uncanny valley: does it exist. In *Proceedings of conference of human computer interaction, workshop on human animated character interaction*.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British journal of psychology*, *77*(3), 305-327.

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

Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., & Jancke, L. (2013). Category processing and the human likeness dimension of the uncanny valley hypothesis: eye-tracking data. *Frontiers in psychology*, *4*, 108.

Cunningham, M. R. (1986). Measuring the physical in physical attractiveness: quasi-experiments on the sociobiology of female facial beauty. *Journal of personality and social psychology*, *50*(5), 925.

Cunningham, M.R., Barbee A.P., & Philhower C. (2002). Dimensions of facial physical attractiveness: The intersection of biology and culture, in *Facial Attractiveness: Evolutionary, Cognitive, and Social Perspectives*. Westport, Conn.: Ablex Publishing.

Currie, T. E., & Little, A. C. (2009). The relative importance of the face and body in judgments of human physical attractiveness. Evolution and Human Behavior, 30(6), 409–416. https://doi.org/10.1016/j.evolhumbehav.2009.06.005

Dietsch, J. (2010). People meeting robots in the workplace [industrial activities]. *IEEE Robotics & Automation Magazine*, *17*(2), 15-16.

Eimer, M., & Holmes, A. (2002). An ERP study on the time course of emotional face processing. *Neuroreport*, *13*(4), 427-431.

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychological review*, *116*(4), 752.

Gothard, K. M., Brooks, K. N., & Peterson, M. A. (2009). Multiple perceptual strategies used by macaque monkeys for face recognition. *Animal cognition*, *12*(1), 155-167.

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.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annu. Rev. Neurosci.*, *27*, 649-677.

Hanson, D. (2006). Exploring the aesthetic range for humanoid robots. In *Proceedings of the ICCS/CogSci-2006 long symposium: Toward social mechanisms of android science* (pp. 39-42). Citeseer.

Hanson, D., Olney, A., Prilliman, S., Mathews, E., Zielke, M., Hammons, D., ... & Stephanou, H. (2005). Upending the uncanny valley. In *AAAI* (Vol. 5, pp. 1728-1729).

Haeske, A. B., & Schmettow, M. (2016). *The uncanny valley: involvement of fast and slow evaluation systems* (Bachelor's thesis, University of Twente).

Ho, C. C., & MacDorman, K. F. (2017). Measuring the uncanny valley effect. *International Journal of Social Robotics*, *9*(1), 129-139.

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

Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in psychology*, *6*, 390.

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

Koopman, R., & Schmettow, M. (2019) The Uncanny Valley as a universal experience : a replication study using multilevel modelling. http://purl.utwente.nl/essays/77172

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Early visual experience and face processing. *Nature*, *410*(6831), 890-890.

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

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

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.

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, *7*(3), 297-337.

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

Mathur, M. B., Reichling, D. B., Lunardini, F., Geminiani, A., Antonietti, A., Ruijten, P. A., ... & Aczel, B. (2020). Uncanny but not confusing: Multisite study of perceptual category confusion in the Uncanny Valley. *Computers in Human Behavior*, *103*, 21-30.

McTear, M., Callejas, Z., & Griol, D. (2016). Conversational interfaces: devices, wearables, virtual agents, and robots. In *The Conversational Interface* (pp. 283-308). Springer, Cham.

Mitchell, W. J., Szerszen Sr, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception*, *2*(1), 10-12.

Moll, B., & Schmettow, M. (2015). Investigating the origins of the uncanny valley : The effect of presentation time on ratings of uncanniness.

Moore, R. K. (2012). A Bayesian explanation of the 'Uncanny Valley'effect and related psychological phenomena. *Scientific reports*, *2*(1), 1-5.

Mori, M., MacDorman, K. F., & Kageki, N. (1970/2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98-100.

Most recent common ancestor (2021). In Wikipedia. https://en.wikipedia.org/w/index.php?title=Most_recent_common_ancestor&oldid=1013660292 Retrieved 15.06.2021.

Or, C. C. F., & Wilson, H. R. (2010). Face recognition: Are viewpoint and identity processed after face detection?. *Vision research*, *50*(16), 1581-1589.

Papeo, L. (2020). Twos in human visual perception. *Cortex*, *132*, 473-478.

Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., Coleman, M., & Nelson, C. A. (2005). Plasticity of face processing in infancy. *Proceedings of the national academy of sciences*, *102*(14), 5297-5300.

Primate. (2021). In Wikipedia. https://en.wikipedia.org/w/index.php?title=Primate&oldid=1022998884 Retrieved 15.06.2021.

Rádlová, S., Landová, E., & Frynta, D. (2018). Judging others by your own standards: attractiveness of primate faces as seen by human respondents. *Frontiers in psychology*, *9*, 2439.

Schmettow, M. (2021). New Statistics for design researchers. Springer International Publishing. Chapters 5.5-8. https://doi.org/10.1007/978-3-030-46380-9

Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, *69*(3), 243-265.

Seyama, J. I., & 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.

Slijkhuis, P. J., & Schmettow, M. (2017). *The Uncanny Valley Phenomenon: A replication with short presentation times* (Master's thesis, University of Twente).

Steckenfinger, S. A., & Ghazanfar, A. A. (2009). Monkey visual behavior falls into the uncanny valley. *Proceedings of the National Academy of Sciences*, *106*(43), 18362-18366.

Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology*, *19*(4), 393-407.

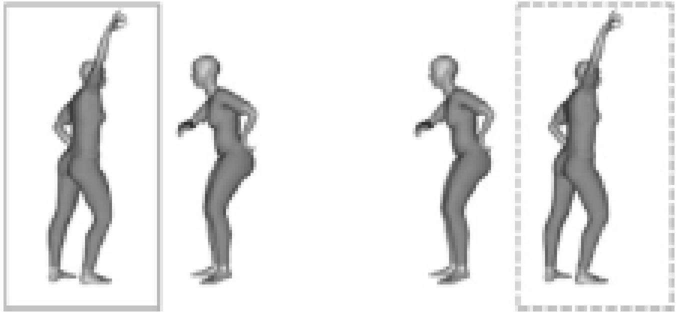
Yu, N. (2001). What Does Our Face Mean to Us? Pragmatics & Cognition, 9, 1–36. https://doi.org/10.1075/pc.9.1.02yu

Zhang, J., Li, S., Zhang, J. Y., Du, F., Qi, Y., & Liu, X. (2020). A Literature Review of the Research on the Uncanny Valley. In *International Conference on Human-Computer Interaction* (pp. 255-268). Springer, Cham.

Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM computing surveys (CSUR)*, *35*(4), 399-458.

Zhao, Y., Zhen, Z., Liu, X., Song, Y., & Liu, J. (2018). The neural network for face recognition: Insights from an fMRI study on developmental prosopagnosia. *NeuroImage*, *169*, 151-161.

Appendix A Stimuli of facing and non-facing body-pairs from the study of Papeo (2020)



Appendix B Consent Form

Information for participation in the research study on humans' emotional response towards different faces

You are being invited to participate in a research study conducted as part of a bachelor thesis within the Faculty of Behavioural, Management and Social Sciences at the University of Twente. The aim of this research is to assess humans' emotional response towards different faces. Completing the study will take approximately 30 minutes and involves viewing images and answering survey questions.

Risks
To our understanding, there are no risks associated with this research study, however, some images might be sensitive or provoke personal discomfort. Given your participation in this study is entirely voluntary, you are free to withdraw at any time without providing any reasons or experiencing any disadvantages.

Handling of data and confidentiality
To the best of our ability, your answers in this study will remain confidential. We will minimize any risks by assuring that your participation in this study is anonymous, as no information will be collected that would allow personal identification. All data will be stored securely and will not be shared with anyone outside the research team. Any comments you make may be quoted in an anonymised form in the research papers resulting from this study.

Contact information
If you have any questions or concerns about this survey or wish for your data to be deleted you can contact the researchers via email:
Lara Geue (l.geue@student.utwente.nl),
Milan Bischoff (m.bischoff@student.utwente.nl),
Marcel Pertenbreiter (m.pertenbreiter@student.utwente.nl),
Jana Westermann (j.m.l.westermann@student.utwente.nl).

If you have questions about your rights as a research participant or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher, please contact the secretary of the Ethics Committee of the Faculty of Behavioural, Management and Social Sciences of the University of Twente (ethicscomittee-bms@utwente.nl).

Consent
I confirm that I have read and understood the information provided above and consent voluntarily to be a participant in this study. I further confirm that I possess a sufficient level of proficiency in English and that I am at least 18 years old.

☐ I consent.
☐ I do not consent to take part in this study.

Appendix C Inclusion and Exclusion criteria for stimuli collection

The following criteria were adopted mainly from the criteria defined by Mathur & Reichling (2016):

Inclusion criteria:

1. All facial features and the outline of the face are discernable.
2. The face is shown in frontal to 3/4 aspect (both eyes visible).
3. The individual belongs to a species that has lived or is still living (no fictitious faces)
4. The image depicts
   a) A living individual
   b) real remainders of the animal that have been taxidermied to closely resemble the way it looked when it was alive, or
   c) A realistic hominid bust (no artificially created CG or morphed face)
5. In case if b) and c), it is shown as it looked when it was alive (not missing any hair, facial features or skin)
6. The resolution of the photo is sufficient to yield a final cropped image of 450x450 pixel.

Exclusion criteria:

1. The individual represents a famous person (for the human image)
2. The image shows other faces or body parts that would appear in the final image.
3. Objects or text overlap the face.

Appendix D Stimuli groups

-AncestralCloseness =

0 = Homo Sapiens

1 = Neanderthalensis

2 = Homo Erectus/Naledi/Ergaster

3 = Hominina (Homo Habilis, Australopithecus afarensis/africanus/sediba, Paranthropus AethiopicusBoisei/Robustus)

4 = Hominini (Panina Pan, Pan troglodytes/paniscus)

5 = Homininae (Gorillini, Gorialla gorilla/beringei)

6 = Hominidae (Ponginae, Pongo abelii/pygmaeus/ tapanuliensis)

7 = Catarrhini (Old world monkeys (Cercopithecidae), Macaca)

8 = Simians/anthropoids/higher order primates (New world monkeys (Platyrrhini), Callitrichdae, Tamarins/Marmosets)

9 = Prosimians/Primates (Strepsirrhini, Lemurs)

10 = Robots

-MorphometricalSimilarity =

1 = 1 (Neanderthalensis) – 3 (Homimina)

2 = 4 (Hominini) – 7 (Catarrhini)

3 = 8 (Simians)

4 = 9 (Prosimians)

Group Ancestral Closeness	0	1	2	3	4	5	6	7	8	9	10
Category	-	1	1	1	2	2	2	2	3	4	-
Number of stimuli	10	3	11	16	16	9	4	11	5	2	11

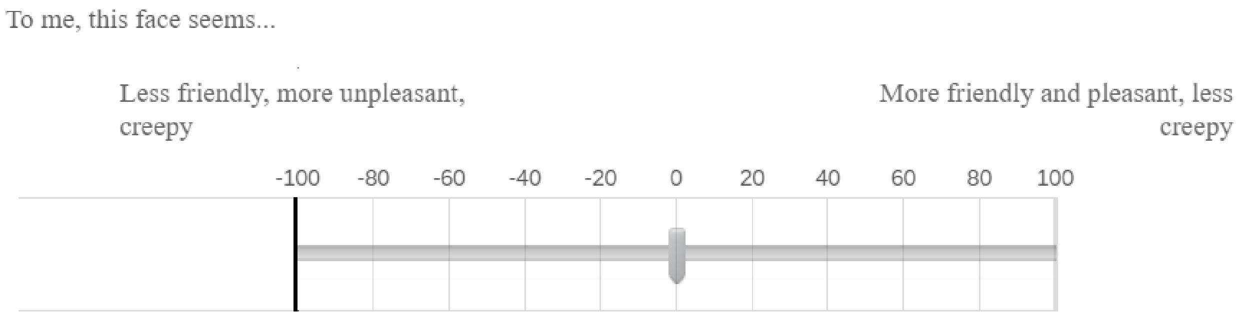
Table 1. Frequency of stimuli for each category

Appendix E Face Rating Scale Translations

Likability scale translations

English	Dutch	German
Less friendly, more unpleasant, creepy	Minder vriendelijk, onaangenaamer, griezelig	Weniger freundlich, unangenehmer, gruselig
More friendly and pleasant, less creepy	Vriendelijker, aangenaamer, minder griezelig	Freundlicher, angenehmer, weniger gruselig

Figure X
Likability VAS with slider in English.



To me, this face seems...

Less friendly, more unpleasant, creepy

More friendly and pleasant, less creepy

Table X

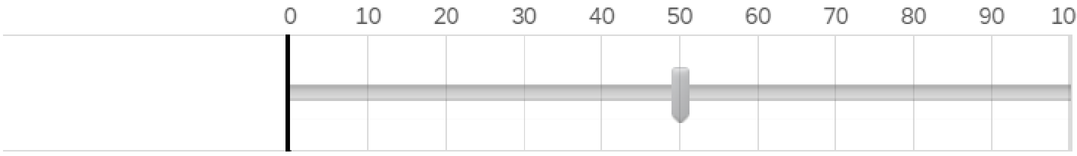
Translations of the eeriness scales

English	Dutch	German
Uninspiring - Spine-tingling	Oninteressant - Opwindend	Uninteressant - Aufregend
Boring - Shocking	Saai - Schokkend	Langweilig - Schockierend
Predictable - Thrilling	Vorspelbaar - Spannend	Vorhersehbar - Spannend
Bland - Uncanny	Flauw - Verontrustend	Fade - Beunruhigend
Unemotional - Hair-raising	Emotieloos - Doodeng	Emotionslos - Haarsträubend

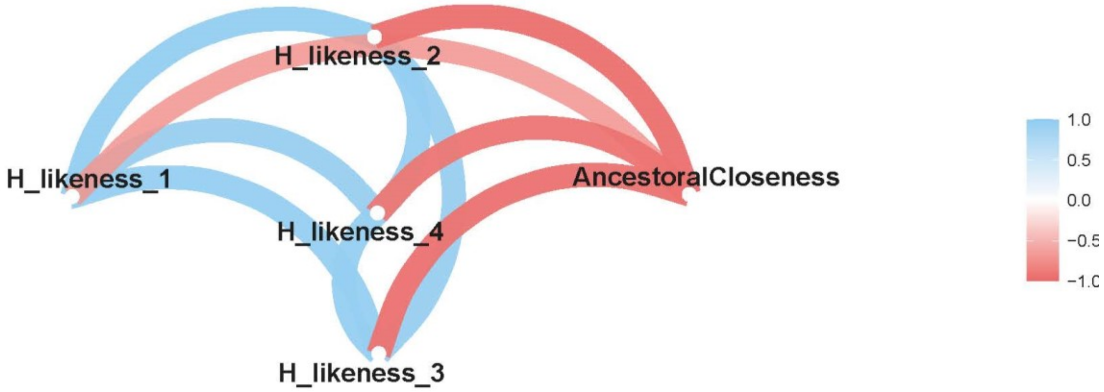
Figure X
Eeriness VAS A) with slider in English.

To me, this face seems...

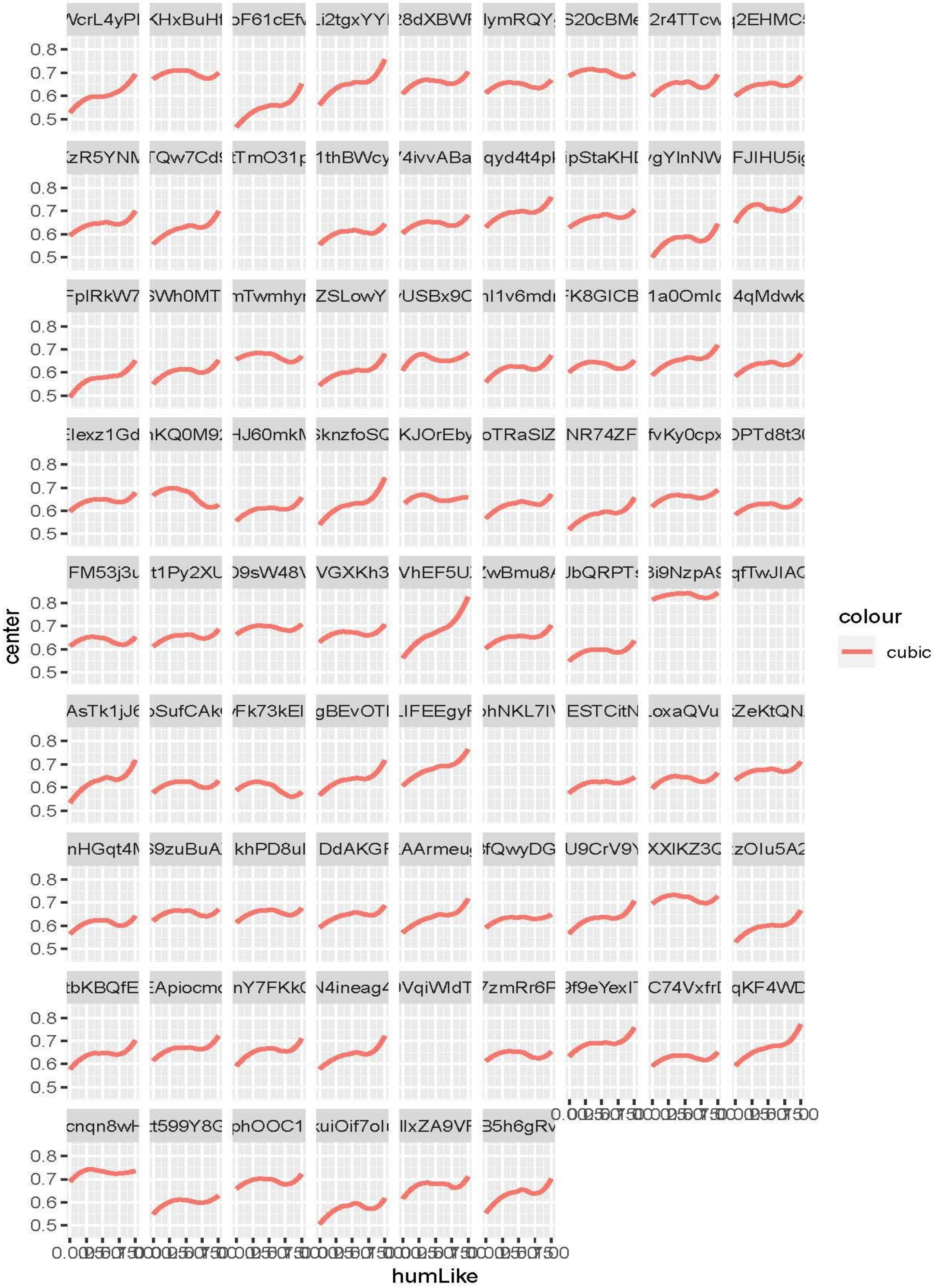
Uninspiring Spine-tingling



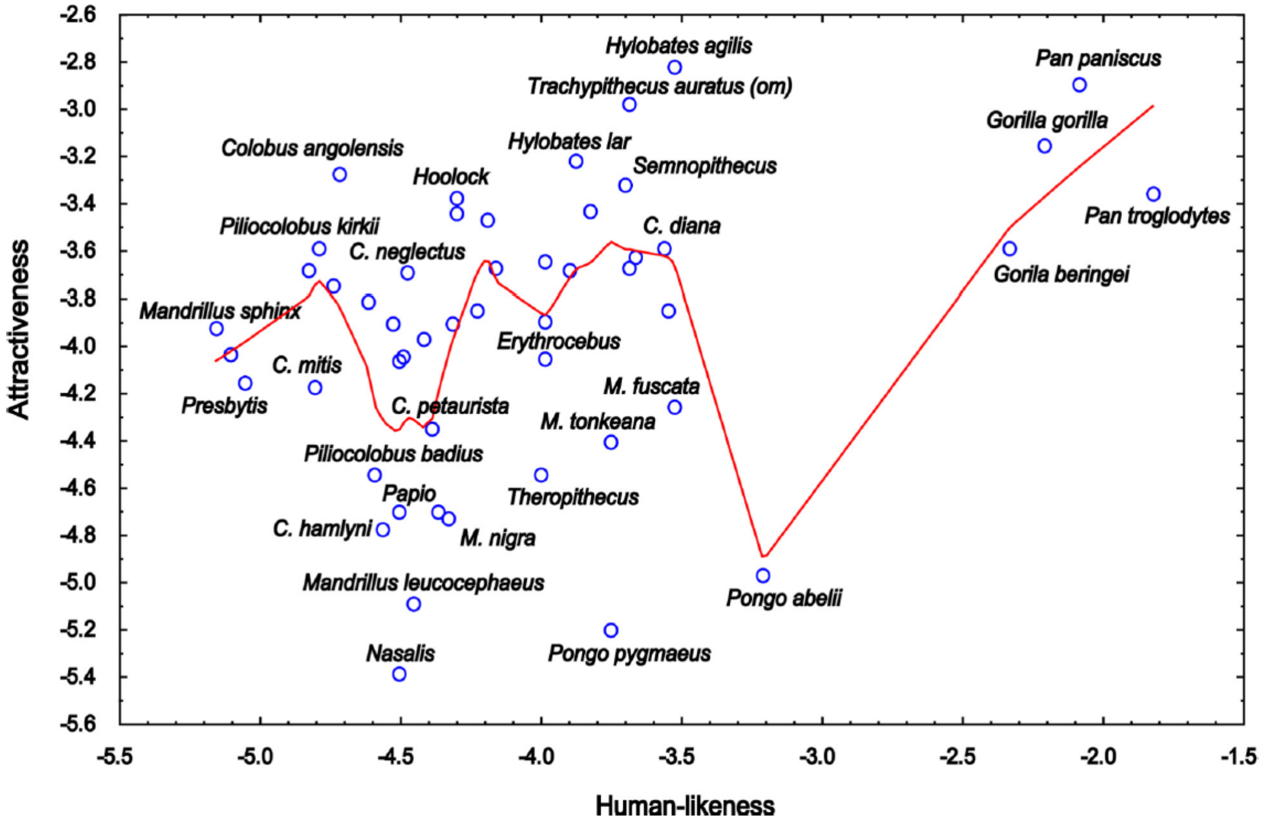
Appendix F Network Plot of correlations between human-likeness ratings and ancestral closeness



Appendix G Individual plots for each participant including the distributional analysis



Appendix F Results from Rádlová, Landová, and Frynta (2018)



Computation of loo-IC as described in 2.6.1 on the dataset of Rádlová, Landová, and Frynta (2018). Based on the IC the linear model describes the data best and consequently the Uncanny Valley is not represented by the data.

Model	IC	Estimate	SE	diff_IC
Linear model (1st degree)	looic	-4.504289	9.855807	0.0000000
Quadratic model (2nd degree)	looic	-3.692093	9.482880	0.8121966
Cubic model (3rd degree)	looic	-1.591048	9.599951	2.9132419
Grand mean (0th degree)	looic	-1.120659	9.218283	3.3836300

Appendix G Data analysis protocol

The following code and output for the data analysis in R was created by Dr. Martin Schmettow (Schmettow, 2021).

Selection of robot faces for comparison

Stimuli from Koopman and Schmettow (2019)

```
load("Uncanny.Rda")
attach(Uncanny)

## The following objects are masked _by_ .GlobalEnv:
##
##     M_poly_2, M_poly_3, P_univ_uncanny

## The following object is masked from package:uncanny:
##
##     trough

predict(M_poly_3) %>%
  left_join(RK_2) %>%
  mutate(Range = round(huMech * 100) %/% 10) %>%
  mutate(error = abs(avg_like - center)) %>%
  group_by(Range) %>%
  mutate(Best = (error == min(error))) %>%
  ungroup() %>%
  filter(Best) %>%
  select(Range, Stimulus, huMech, error) %>%
  arrange(Range)

## Joining, by = "Obs"
```

Range	Stimulus	huMech	error
0	3	0.0375	1.126914
1	15	0.1875	1.101790
2	22	0.2750	1.151271
3	28	0.3500	1.137703
4	32	0.4000	1.176306
5	42	0.5250	1.149822
6	50	0.6250	1.261915
7	59	0.7375	1.236049
8	68	0.8500	1.117708
9	74.2	0.9275	1.129486

```
detach(Uncanny)
```

Stimuli

```
Stimuli <-
  readxl::read_excel("Data/BA21/Stimuli.xlsx") %>%
  mutate(Stimulus = str_c("S", as.character(Stimulus))) %>%
  mutate(Set = factor(Set, 1:4, labels = c("Primate", "Human", "Tri23",
"Robot"))) %>%
  mutate(humLike = rowMeans(select(., starts_with("H_like")), na.rm = T)) %>%
  mutate(valence = rowMeans(select(., starts_with("E_valence")), na.rm = T))

Stimuli %>%
  sample_n(20)
```

Stimulus	Set	H_likeness_1	H_likeness_2	H_likeness_3	H_likeness_4	E_1	E_2	E_3	E_4	E_valence_1	E_valence_2	E_valence_3	E_valence_4	AncestoralCloseness	Ethnicity	humLike	valence
S78	Primate	55	65	65	55	1	4	4	1	65	80	40	70	4	0	60.00	63.75
S25	Primate	97	96	93	95	1	1	1	1	37	60	20	25	1	0	95.25	35.50
S72	Primate	82	75	65	70	0	0	0	0	-10	0	0	0	3	0	73.00	-2.50
S40	Primate	18	50	45	50	4	4	4	4	10	-20	-20	5	5	0	40.75	-6.25
S57	Primate	40	50	55	50	0	0	0	0	0	0	0	5	4	0	48.75	1.25
S84	Primate	97	90	94	90	5	5	5	4	80	-100	-95	-60	2	0	92.75	-43.75
S38	Robot	37	NA	NA	NA	4	4	4	4	30	0	-20	-15	10	0	37.00	-1.25
S54	Primate	2	10	15	5	5	0	0	0	-26	0	0	0	9	0	8.00	-6.50
S98	Hu	100	100	100	100	4	4	2	4	85	0	-95	-70	0	1	100.0	-20

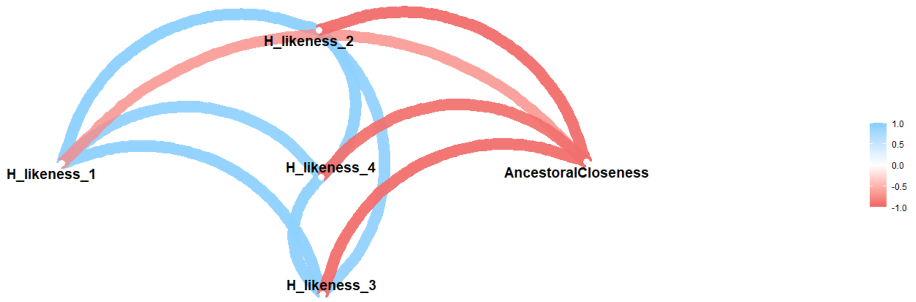
	man															0	.00
S42	Primate	94	90	90	85	1	1	1	4	33	50	30	5	2	0	89.75	29.50
S85	Primate	91	90	95	95	4	4	2	2	-25	-30	-90	-50	2	0	92.75	-48.75
S49	Tri23	100	100	98	97	1	1	1	1	60	70	50	10	0	0	98.75	47.50
S53	Primate	83	90	89	80	0	1	1	1	13	60	40	5	2	0	85.50	29.50
S76	Primate	80	80	82	80	0	5	0	0	0	-50	0	0	3	0	80.50	-12.50
S61	Primate	1	20	15	3	2	2	0	0	-20	-10	0	0	8	0	9.75	-7.50
S93	Human	100	100	100	100	5	6	6	6	-85	-60	-90	-70	0	1	100.00	-76.25
S12	Primate	24	45	25	25	4	4	2	4	70	-20	-80	-70	7	0	29.75	-25.00
S69	Primate	28	45	35	30	4	4	4	4	55	20	10	25	7	0	34.50	27.50
S43	Primate	45	50	60	40	0	3	0	0	-10	-50	0	0	4	0	48.75	-15.00
S37	Primate	70	85	80	70	6	6	1	1	-30	-20	90	20	3	0	76.25	15.00

Correlating ancestral closeness and human-likeness

```
Stimuli %>%
  select(starts_with("H_like"), AncestoralCloseness) %>%
  corrr::correlate()

##
## Correlation method: 'pearson'
## Missing treated using: 'pairwise.complete.obs'
```

term	H_likeness_1	H_likeness_2	H_likeness_3	H_likeness_4	AncestoralCloseness
H_likeness_1	NA	0.9554019	0.9436330	0.9328772	-0.7707424
H_likeness_2	0.9554019	NA	0.9303407	0.9426051	-0.9422645
H_likeness_3	0.9436330	0.9303407	NA	0.9438880	-0.9379207
H_likeness_4	0.9328772	0.9426051	0.9438880	NA	-0.9268578
AncestoralCloseness	-0.7707424	-0.9422645	-0.9379207	-0.9268578	NA

```
Stimuli %>%
  select(starts_with("H_like"), AncestoralCloseness) %>%
  corrr::correlate() %>%
  corrr::network_plot()

##
## Correlation method: 'pearson'
## Missing treated using: 'pairwise.complete.obs'
```



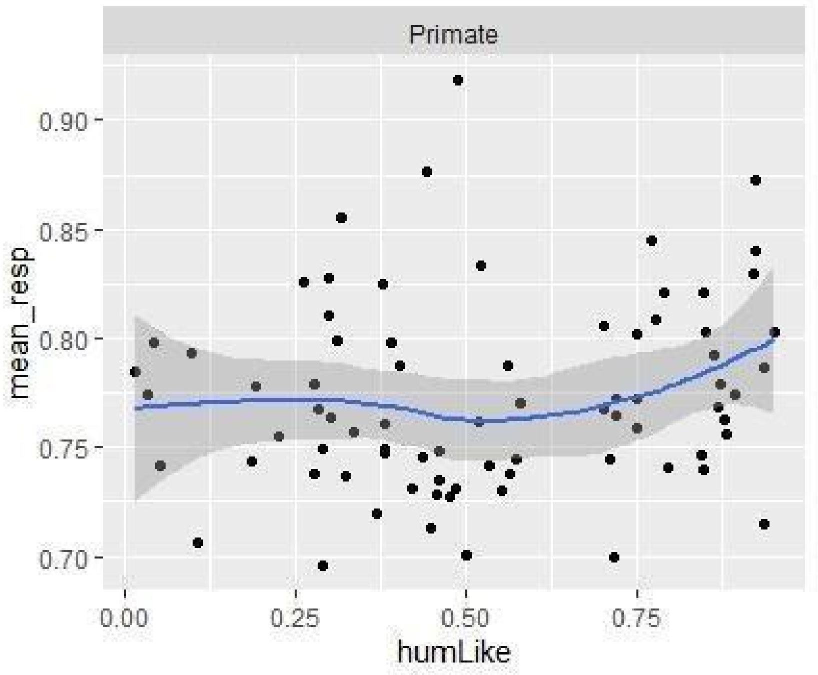
Reading and preparing data

```
D_raw <-
  readxl::read_excel("Data/BA21/Final Dataset Uncanny Valley 16-05-21.xlsx")
%>%
  filter(StartDate != "Start Date")

BA21 <- D_raw %>%
```

```
  select(Part = ResponseId, matches("^S\\d+")) %>%
  pivot_longer(-Part, names_to = "Trial", values_to = "response") %>%
  filter(!is.na(response)) %>%
  separate(Trial, into = c("Stimulus", "Item", "attempt")) %>%
  left_join(Stimuli %>% select(Stimulus, Set, valence, humLike), by =
"Stimulus") %>%
  mutate(Scale = if_else(str_detect(Item, "^2"),
                         "Eeriness",
                         "Display"),
         response  = mascutils::rescale_unit(as.numeric(response)),
         response  = mascutils::rescale_centered(response, scale = .999),
         humLike   = mascutils::rescale_unit(humLike),
         #Stimulus = str_extract(Trial, "^S\\d{2,3}_\\d"),
         #Item     = str_extract(Trial, ),
         #humLike  = as.numeric(str_extract(Stimulus, "\\d{1,3}")),
         humLike_2 = humLike^2,
         humLike_3 = humLike^3) %>%
  bayr::as_tbl_obs()
BA21

BA21_prim <-   BA21 %>%
  filter(Set == "Primate")

BA21 %>%
  group_by(Part) %>%
  summarize(N = n())

BA21 %>%
  group_by(Stimulus) %>%
  summarize(N = n())

BA21_agg <-
  BA21_prim %>%
  filter(Scale == "Eeriness") %>%
  group_by(Stimulus, Set, valence, humLike) %>%
  summarize(mean_resp = mean(response, na.rm = T)) %>%
  ungroup() %>%
  as_tbl_obs()


BA21_agg
```

For a first look at the data, we average the eeriness responses over the stimuli so that every stimulus is assigned an eeriness score.

```
BA21_agg %>%
  ggplot(aes(x = humLike,
             y = mean_resp,
             size = valence)) +
  geom_point() +
  geom_smooth() +
  facet_wrap(~Set)

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Population-level model

We estimate four polynomials, namely grand mean, linear, quadratic and cubic and include emotional valence as a control variable.

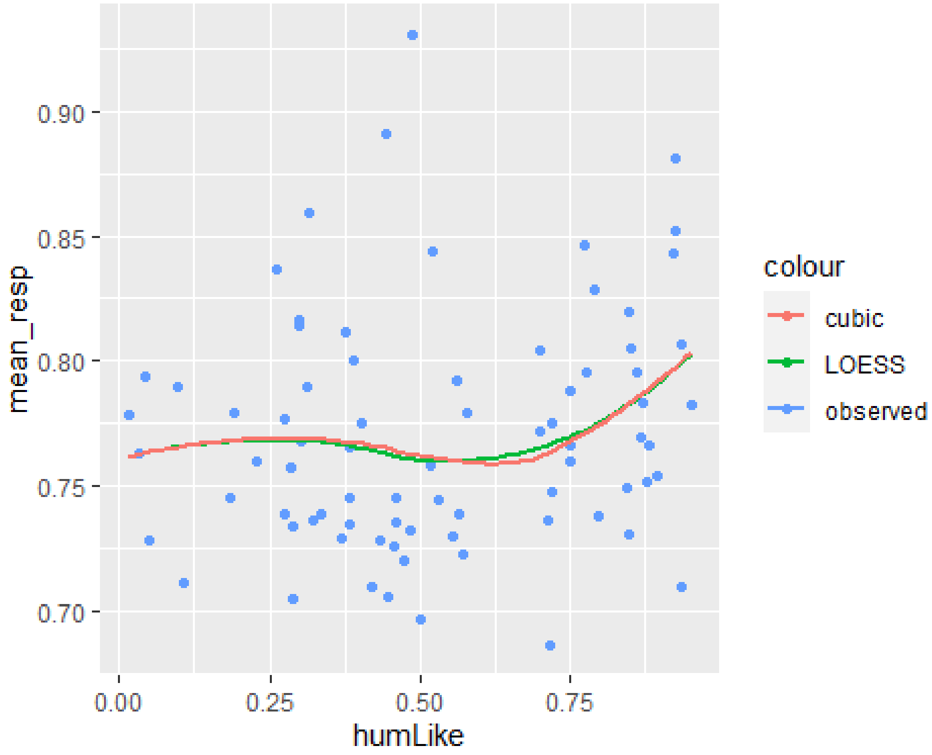
```
M_poly_3 <- stan_glm(mean_resp ~ valence + poly(humLike, 3),
                data = BA21_agg)

M_poly_2 <- stan_glm(mean_resp ~ valence + poly(humLike, 2),
                data = BA21_agg)

M_poly_1 <- stan_glm(mean_resp ~ valence + poly(humLike, 1),
                data = BA21_agg)

M_poly_0 <- stan_glm(mean_resp ~ 1 + valence,
                data = BA21_agg)

clu(M_poly_3)
```

Parameter estimates with 95% credibility limits

parameter	fixef	center	lower	upper
Intercept	Intercept	0.5760378	0.5110060	0.6426081
poly(hum_like, 3)1	poly(hum_like, 3)1	0.5122479	0.0501577	0.9618364
poly(hum_like, 3)2	poly(hum_like, 3)2	0.2039400	-0.2525294	0.6671140
poly(hum_like, 3)3	poly(hum_like, 3)3	-0.0764050	-0.4945528	0.3797763
sigma_resid	NA	0.2275051	0.1880630	0.2830139

```
PP_poly_3 <- post_pred(M_poly_3)

PP_poly_3 %>%
  predict() %>%
  left_join(BA21_agg, by = "Obs") %>%
```

```
  ggplot(aes(x = humLike)) +
  geom_point(aes(y = mean_resp, color = "observed")) +
  geom_smooth(aes(y = mean_resp, color = "LOESS"), se = F) +
  geom_smooth(aes(y = center, color = "cubic"), se = F)

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Model selection

We use the leave-one-out (LOO) approximation to acquire and compare the relative predictive accuracy of all four population-level models.

```
Loo_poly <-
  list(loo(M_poly_0),
       loo(M_poly_1),
       loo(M_poly_2),
       loo(M_poly_3))

compare_IC(Loo_poly)
```

Polynomials ranked by their predictive accuracy (LOO-IC)

Model	IC	Estimate	SE	Diff_IC

M_poly_1	looic	-268.6983	12.24563	0.0000000
M_poly_0	looic	-268.1633	12.19914	0.5350387
M_poly_2	looic	-266.5040	13.08301	2.1943270
M_poly_3	looic	-264.3055	13.12711	4.3928425

We can see that the cubic model is not the preferred one on population level, as one might expect for the uncanny valley effect. However, in the later steps of analysis we will observe that individual differences distort the graph on population level. Therefore, we use the cubic model to create a test statistic on the MCMC samples to give us the probability of obtaining an uncanny valley curve on population level.

```
P_wide <-
  posterior(M_poly_3) %>%
  as_tibble() %>%
  filter(type == "fixef") %>%
  select(chain, iter, fixef, value) %>%
  pivot_wider(id_cols = c("chain", "iter"),
              names_from = fixef,
              values_from = value) %>%
  mutate(shoulder = uncanny::shoulder(.[3:6]),
         trough = uncanny::trough(.[3:6]),
         is_uncanny = !is.na(shoulder) & !is.na(trough))

cat("The probability of the population level being a UV curve is: ",
mean(P_wide$is_uncanny))

## The probability of the population level being a UV curve is:  0.6315
```

Multilevel model

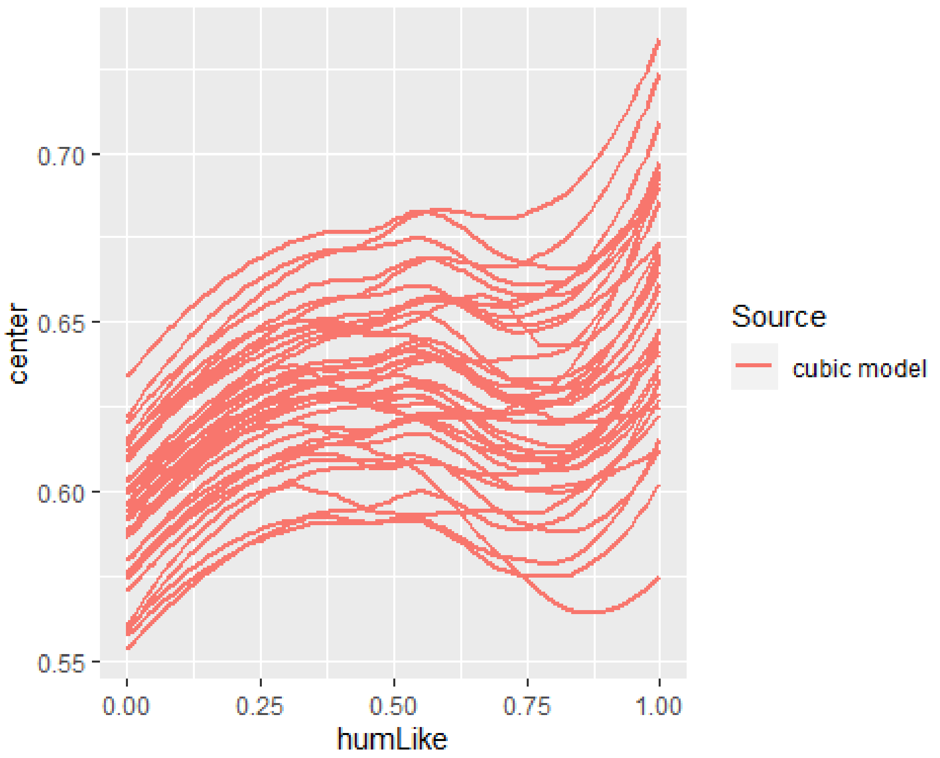
```
M_poly_4 <-
  stan_glmer(response ~ 1 + valence + humLike + humLike_2 + humLike_3 +
             (1 + valence + humLike + humLike_2 + humLike_3|Part),
             data = BA21)

PP_poly_4 <- post_pred(M_poly_4, thin = 5)

fixef_ml(M_poly_4)
```

Population-level coefficients with random effects standard deviations

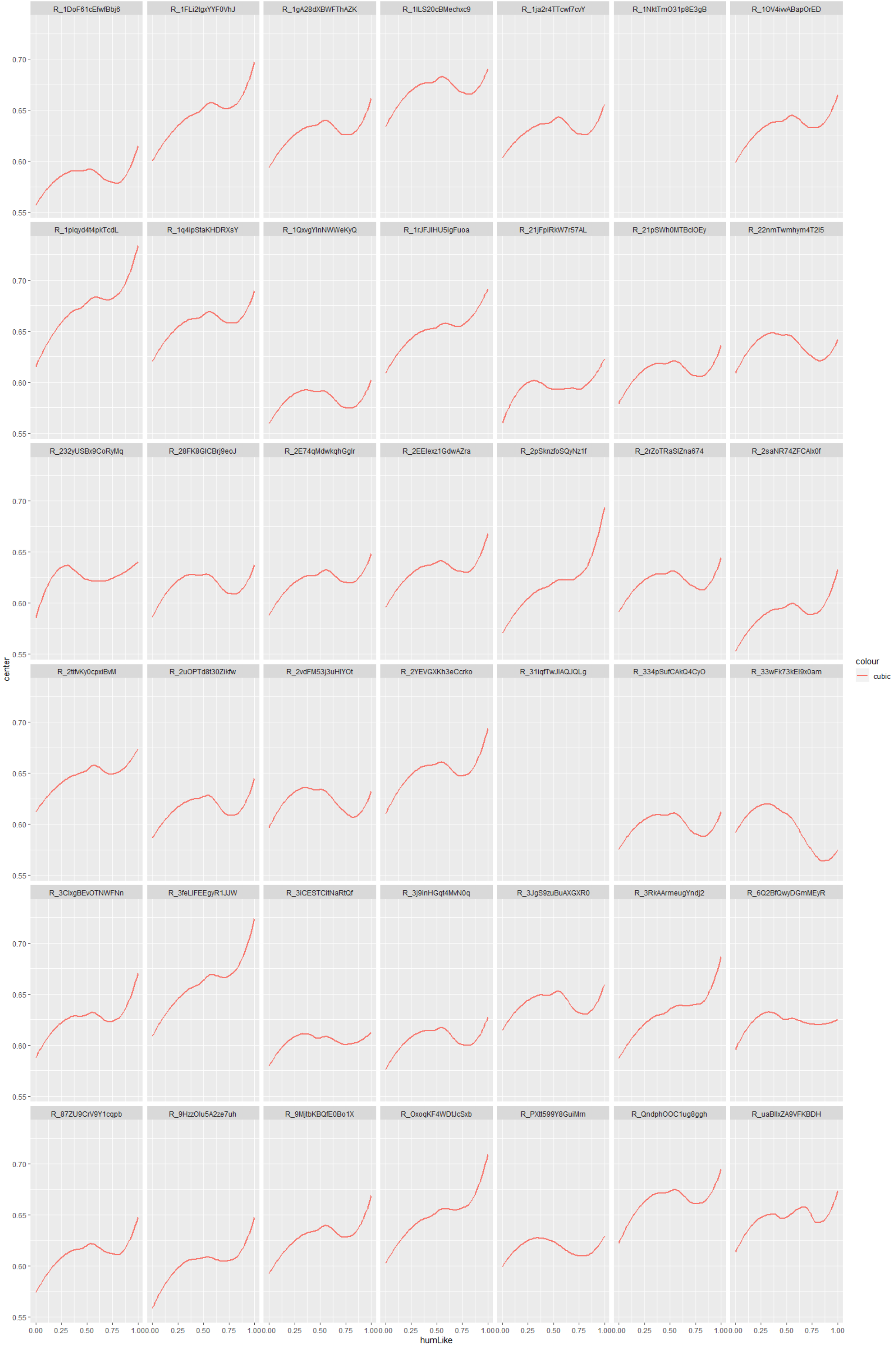
fixef	center	lower	upper	SD_Part
Intercept	0.5833803	0.5583426	0.6080007	0.0265205
valence	0.0011061	0.0009101	0.0012908	0.0002134
humLike	0.4778191	0.3051359	0.6479144	0.0179322

humLike_2	-1.1130629	-1.4843434	-0.7271630	0.0235903
humLike_3	0.7180369	0.4709761	0.9552268	0.0236564

```
PP_poly_4 %>%
  predict() %>%
  left_join(BA21, by = "Obs") %>%
  ggplot(aes(x = humLike)) +
  geom_smooth(aes(y = center, color = "cubic model",
                  group = Part), se = F) +
  labs(color = "Source")

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



```
PP_poly_4 %>%
  predict() %>%
  left_join(BA21, by = "Obs") %>%
  ggplot(aes(x = humLike)) +
  facet_wrap(~Part) +
  #geom_point(aes(y = response, color = "observed")) +
  #geom_smooth(aes(y = response, color = "LOESS"), se = F) +
  geom_smooth(aes(y = center, color = "cubic",
                  group = Part), se = F)

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



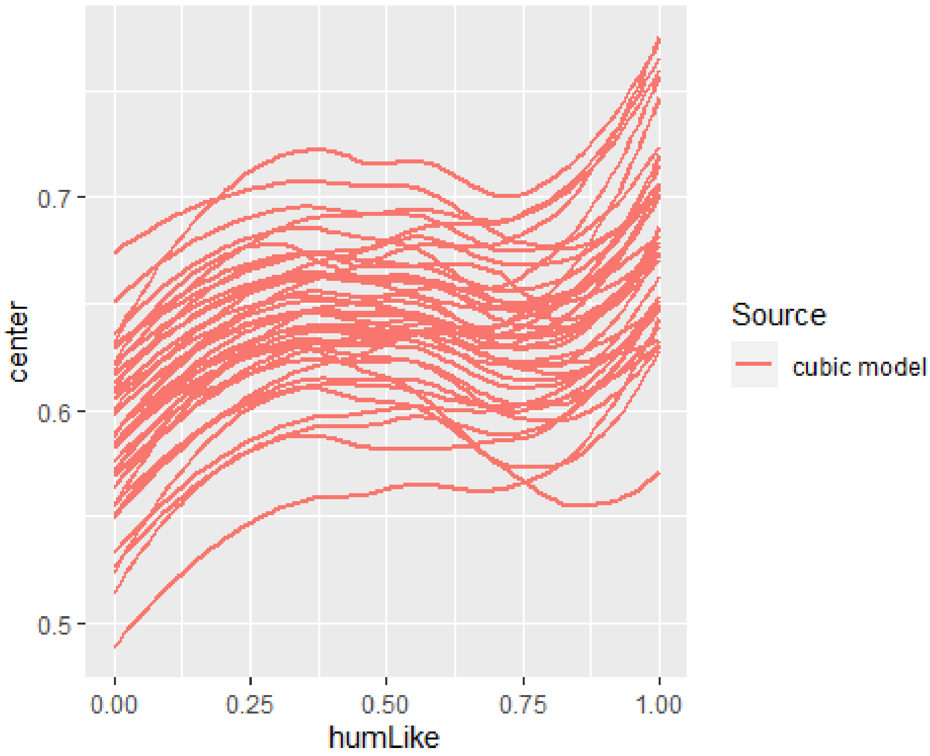
Distributional Beta regression

```r
M_poly_5 <-
  brm(bf(response ~ 1 + valence + humLike + humLike_2 + humLike_3 +
           (1 + valence + humLike + humLike_2 + humLike_3|Part),
                    phi ~ 1 + (1 | Part)),
             family = Beta(),
             data = BA21,
      inits = "0")

PP_poly_5 <- post_pred(M_poly_5, thin = 5)

P_poly <- posterior(M_poly_5)

fixef_ml(M_poly_5)
```

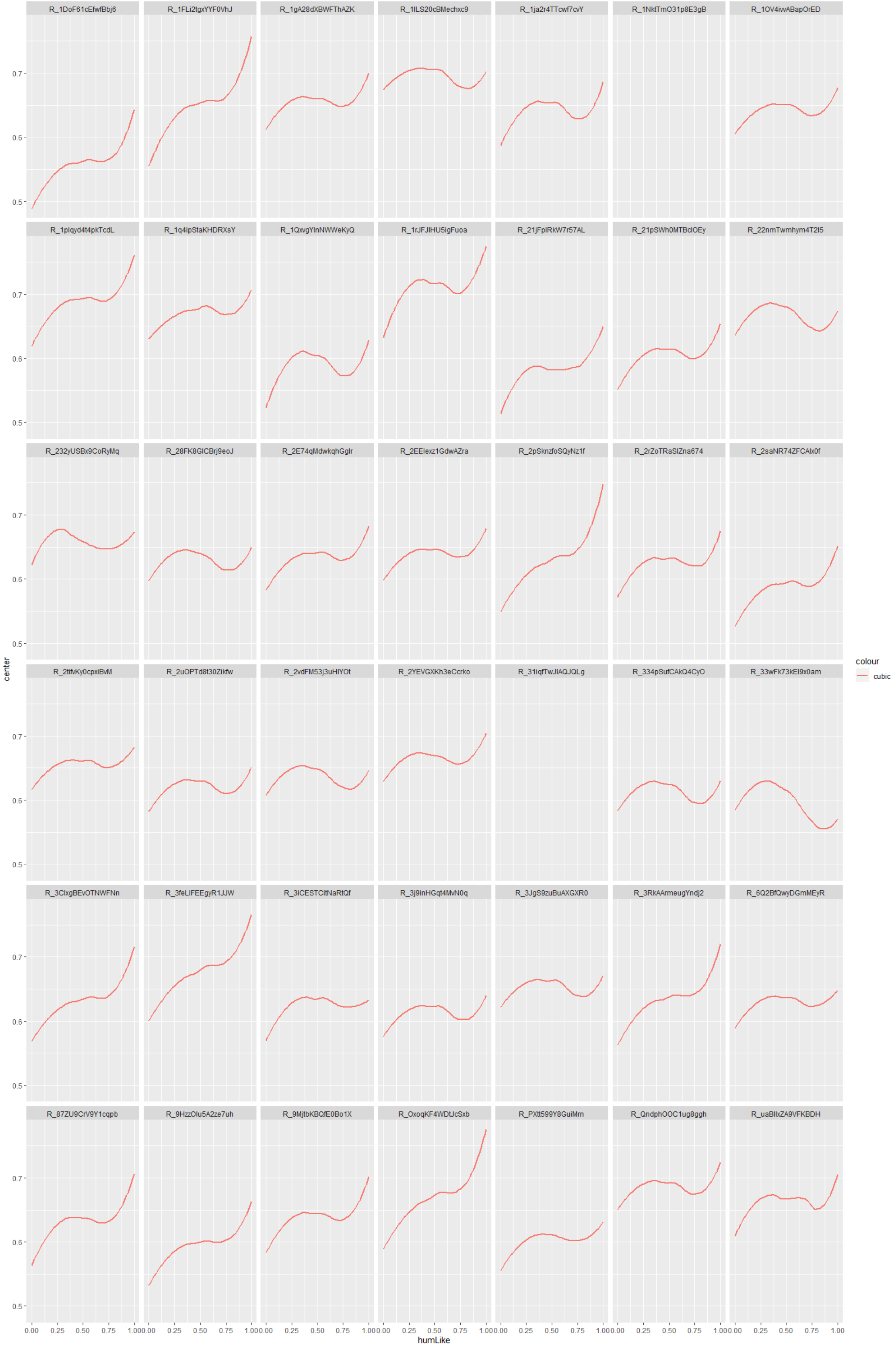
Population-level coefficients with random effects standard deviations

fixef	center	lower	upper	SD_Part	
Intercept	0.2423457	0.1349185	0.3537239	0.1725790	NA
NA	1.2259843	1.0541065	1.4012498	NA	0.5545595
valence	0.0035227	0.0026571	0.0044081	0.0012479	NA
humLike	2.1916328	1.4888111	2.8686954	0.1213779	NA
humLike_2	-5.0889238	-6.6040510	-3.5435657	0.1514233	NA
humLike_3	3.3125477	2.3226758	4.2920712	0.1318868	NA

```r
PP_poly_5 %>%
  predict() %>%
  left_join(BA21, by = "Obs") %>%
  ggplot(aes(x = humLike)) +
  geom_smooth(aes(y = center, color = "cubic model",
                  group = Part), se = F) +
  labs(color = "Source")

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



```
T_predicted <-
  PP_poly_5 %>%
  predict() %>%
  right_join(BA21, by = "Obs")


T_predicted %>%
  ggplot(aes(x = humLike)) +
  facet_wrap(~Part) +
  #geom_point(aes(y = response, color = "observed")) +
  #geom_smooth(aes(y = response, color = "LOESS"), se = F) +
  geom_smooth(aes(y = center, color = "cubic",
                  group = Part), se = F)

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



```r
write_csv(T_predicted, file = "T_predicted.csv")
```

Universality

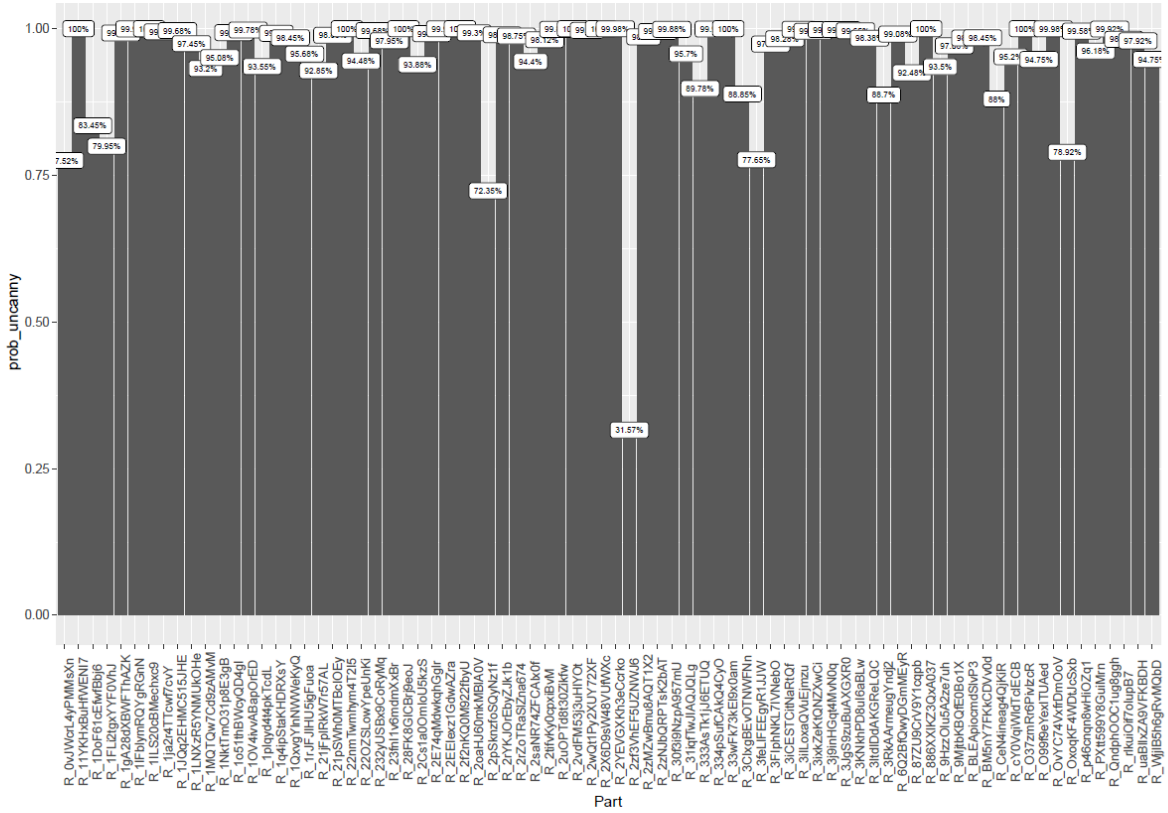
```
P_univ_uncanny <-
  posterior(M_poly_5) %>%
  #as_tibble() %>%
  dplyr::filter(is.na(nonlin), fixef != "valence") %>%
  re_scores() %>%
  select(iter, Part = re_entity, fixef, value) %>%
  tidyr::spread(key = "fixef", value = "value") %>%
  select(iter, Part,
         humLike_0 = Intercept,
         humLike_1 = humLike, humLike_2, humLike_3) %>%
  mutate(
    trough = trough(select(., humLike_0:humLike_3)),
    shoulder = shoulder(select(., humLike_0:humLike_3)),
    has_trough = !is.na(trough),
    has_shoulder = !is.na(shoulder),
    shoulder_left = trough > shoulder,
    is_uncanny = has_trough & has_shoulder & shoulder_left
  )

P_univ_uncanny %>%
  select(Part:shoulder) %>%
  pivot_longer(humLike_0:shoulder,
               names_to = "parameter",
               values_to = "value") %>%
  group_by(Part, parameter) %>%
  summarize(center = median(value, na.rm = T)) %>%
  pivot_wider(names_from = "parameter",
              values_from = "center") %>%
  write_csv(file = "Participant_level.csv")
```

Individual positions of shoulder and trough

```
P_univ_uncanny %>%
  ggplot(aes(x = Part)) +
  geom_violin(aes(y = trough, color = "Trough")) +
  geom_violin(aes(y = shoulder, color = "Shoulder")) +
#  theme(axis.text.x = element_text(angle = 90)) +
  coord_flip()
```

```
P_univ_uncanny %>%
  group_by(Part) %>%
  summarize(prob_uncanny = mean(is_uncanny, na.rm = T)) %>%
#  ungroup() %>%
#  mutate(Part_ord = rank(prob_uncanny)) %>%
  mutate(label = str_c(100 * round(prob_uncanny, 4), "%")) %>%
```

```
  ggplot(aes(x = Part, y = prob_uncanny)) +
  geom_col() +
  geom_label(aes(label = label), size = 2) +
  theme(axis.text.x = element_text(angle = 90))
```



```
save(BA21, BA21_agg, PP_poly_4, M_poly_4, P_univ_uncanny, M_poly_3,
     PP_poly_3, M_poly_2, M_poly_1, M_poly_0, Loo_poly, M_poly_5, PP_poly_5,
     P_univ_uncanny, M_6, M_7,
     file = "Bachelor thesis Uncanny Valley.Rda")
```