At the fringes of normality – a neurocognitive model of the uncanny valley on the detection and negative evaluation of deviations

A dissertation thesis submitted for the degree of Doctor of Philosophy (PhD)

Cardiff University School of Psychology

July 2023

Alexander Diel

Summary

Information violating preconceived patterns tend to be disliked. The term “uncanny valley” is used to described such negative reactions towards near humanlike artificial agents as a nonlinear function of human likeness and likability. My work proposes and investigates a new neurocognitive theory of the uncanny valley and uncanniness effects within various categories. According to this refined theory of the uncanny valley, the degree of perceptual specialization increases the sensitivity to anomalies or deviations in a stimulus, which leads to a greater relative negative evaluation. As perceptual specialization is observed for many human-related stimuli (e.g., faces, voices, bodies, biological motion) attempts to replicate artificial human entities may lead to design errors which would be especially apparent due to a higher level of specialization, leading to the uncanny valley. The refined theory is established and investigated throughout 10 chapters. In Chapters 2 to 4, the correlative (Chapters 2 and 3) and causal (Chapter 4) association between perceptual specialization, sensitivity to deviations, and uncanniness are observed. In Chapters 5 to 6, the refined theory is applied to inanimate object categories to validate its relevance in stimulus categories beyond those associated with the uncanny valley, specifically written text (Chapter 5) and physical places (Chapter 6). Chapters 7 to 10 critically investigate multiple explanations on the uncanny valley, including the refined theory. Chapter 11 applies the refined theory onto ecologically valid stimuli of the uncanny valley, namely an android’s dynamic emotional expressions. Finally, Chapter 12 summarized and discusses the findings and evaluates the refined theory of the uncanny based on its advantages and disadvantages. With this work, I hope to present substantial arguments for an alternative, refined theory of the uncanny that can more accurately explain a wider range of observation compared to the uncanny valley.

Table of Contents

Acknowledgements

I express my greatest gratitude to my supervisor Dr. Michael Lewis, who supported me throughout my PhD years and rekindled in me a spark of curiosity and passion for science that I thought was lost. I thank him for giving me a place to do my research, for his constant encouragement, and for his amazing and steady guidance and support on every step of my PhD journey. I am also extremely grateful to Dr. Karl MacDorman, who ignited my interest in the uncanny valley and who supported me throughout my master's and PhD to establish myself as a young researcher. It is a privilege to stand on both their shoulders.

I express special thanks the collaborators I had throughout my PhD, who offered me amazing opportunities, interesting ideas, and constructive feedback on our work: Specifically, I thank Dr. Takashi Minato, Dr. Wataru Sato, and Dr. Chun-Ting Hsu for our collaboration and my time at the Guardian Robot Project RIKEN in Japan – of course, I also thank RIKEN's android Nikola for being a part of my work. Finally, I thank Dr. Sebastian Ocklenburg, Dr. Malte Elson, and Dr. Denny Han, without whom I could have not completed my EEG work.

My dissertation could not have been achieved without the support of the Studienstiftung des Deutschen Volkes (German Academic Scholarship Foundation) who supported me during both my time in the United Kingdom and Japan. I thank the Foundation for what it allowed me to experience and achieve.

Many thanks to those who stood by me during my PhD, who inspired my ideas, who reminded me to laugh, and who supported me through darker times: I thank Mohammed Hasan, Astrid Hönekopp, Josephine Boegel, Marie von Rogal, Stanislaw Diel, Stefanie May, and Rié Yamaguchi. Of course, I also thank my non-human supporters: Kitrina, Lune, Romana, Tipica, Hekate, Yadwiga, Abigail, and Lili – may the departed rest in peace.

Finally, I dedicate this dissertation to my family, who suffered injustice and horrors, deportation and famine, loss of home and people, and so much more. You, my family, who survived, fought and overcame, and who gave so much to me for a life you never were allowed to live – a gift I always have and always will cherish from the depths my heart. It is the story of our past that taught me the value of life itself. May the ashes of our past nourish a brighter future; may my life and work tip the scales of Justicia and stir the wheels of Fortuna towards the light of hope; and may our future be as fertile and as golden as the fields of your former Volga home.

Chapter 1: The Uncanny Valley

"The "uncanny" is that class of the terrifying which leads back to something long known to us, once very familiar." – Sigmund Freud, The Uncanny (1919)

People prefer to live in predictable environments (Chetverikov & Kristjánsson, 2016; Kaplan & Kaplan, 2011). Prior experience – or familiarity – is one of the major sources to infer predictions about the future. Hence, familiarity fosters a sense of security and confidence in the ability to safely interact with one's surroundings (de Vries, Holland, Starr, & Winkielman, 2010). The preference for familiarity is long-established: Even simple patterns are preferred over atypical, deviating, or novel ones (Gollwitzer, Martel, Heinecke, & Bargh, 2017; Winkielman, Halberstadt, Fazendeiro, & Catty, 2006; Zajonc, 1968). Meanwhile, information that does not fit our neat und usual structures is often described as strange, weird, creepy, eerie, uncanny, or uncomfortable (Burleigh & Schoenherr, 2015; Diel, Weigelt, & MacDorman, 2022; Gollwitzer et al., 2017; Mangan, 2015). Examples of the devaluation of deviating information range from simple patterns (Gollwitzer et al., 2017; Winkielman et al., 2006) to themes of estrangement in literature (Fisher, 2016; Royle, 2003), unusual social situations (Langer & König, 2018), inappropriate or unpredictable human behaviour (McAndrew & Koehnke, 2016), and new technology (Mori, 1970; Tene & Polotensky, 2014). In the following, various examples of disturbing feelings caused by information deviating from familiar patterns are explored and summarized.

Early literature: The Uncanny

Explorations of the disturbing feeling (here: *uncanniness*) caused by deviating stimuli range back to last century's psychoanalysis and psychiatry. According to Jentsch (1907), stimuli

combining new with familiar information, such as realistic dolls or wax figures, are perceived as uncanny due to the perceiver's inability to explain or conceptualize those stimuli, leading to a decrease of intellectual certainty. Meanwhile, Freud (1918) attributed uncanniness to a multitude of psychoanalytic mechanisms, most notably to reminders of something familiar yet repressed. In literature analysis, "the Uncanny" as an aesthetic concept has been associated with estrangement and defamiliarization, the absence of something that should be present or vice versa, the perception of something strange within an intimate environment, or alienation (Fisher, 2016; Royle, 2003; Masschelein, 2011).

While such early explorations of the concept of "the uncanny" lack empirical support, they provide insight into the semantic understanding of the concept. A general theme across these depictions an uncanny stimulus' proximity to something intimate or familiar, yet the stimulus is somehow corrupted (e.g., is missing certain features), repressed, or otherwise made strange.

Familiar and strange: Low-level processing

Effects of familiarity and deviancy on a stimulus' likability are found in simple patterns. Seminal work by Zajonc (1968) coined the term *mere exposure effect* to describe a preference for otherwise novel stimuli that have been exposed frequently compared to less frequent stimuli. Furthermore, a *prototypicality effect* is found for a set of different stimuli considered part of a group or category: Averaged prototypical dot patterns are preferred over more distinct ones (Winkielman et al., 2006), a process thought to be caused by processing disfluency (Reber, Schwarz, & Winkielman, 2004). Finally, *deviancy aversion* is the negative evaluation in simple deviations in low-level patterns (Gollwitzer et al., 2017), and individual differences in deviancy aversion predict negative evaluations of social norms and dislike of statistical minorities (Gollwitzer, Marshall, & Bargh, 2020; Gollwitzer, Martel, Heinecke, & Bargh, 2022), indicating domain-general transferability of aversion towards deviancy.

Deviating human appearance and behaviour

Atypical and disfigured faces

Human appearance varies across a multitude of dimensions, with some individual appearances being closer to a population-based centre than others. In faces, those closer to the average are perceived as more attractive (Langlois & Roggman, 1990). Meanwhile, indicators of face ugliness seem consistent across cultures (Sorokowski & Kościński, 2013), and anomalous or disfigured faces are rated more negatively (Hartung et al., 2019; Jamrozik, Oraa Ali, Sarwer, & Chatterjee, 2019; Stone, 2022; Workman et al., 2021; Zebrowitz & Rhodes, 2004). The categorization of faces as disfigured is associated with disgust sensitivity, indicating that dislike of disfigured faces may be linked to disease avoidance mechanisms (Stone, 2021; Zebrowitz & Rhodes, 2004). In any case, unusual or anomalous characteristics in a face tend to be aesthetically devalued.

Physical disability

Social exclusion of individuals with mental or physical disabilities remains a pressing social issue (Morris, 2001; Kitchin, 1998). Explicit and implicit negative attitudes towards people with disabilities remains a consistent issue throughout literature (Wilson & Scior, 2014), and are present in children (Cameron & Rutland, 2006). Goffman (1963) described three types of sources of stigmatizations, namely 1) group identities such as ethnicity, 2) body abominations, and 3) flawed character traits. Traits associated to body abominations elicit behavioral aversive responses as well as emotions like fear and disgust (Jones et al., 1984). Because implicit negative attitudes towards disabled individuals or organisms are 1) cross-cultural (Wilson & Scior, 2014), 2) present in childhood (Cameron & Rutland, 2006), and 3) found in animals (Packer & Pusey, 1984), some researchers have suggested evolutionary mechanisms of social exclusion, stigmatization, negative attitudes, and avoidance of people with physical disabilities. Theories range from avoidance of poor social exchange partners to

low genetic fitness, avoidance of individuals systematically devalued by the collective, maintenance of positive self and group image, and avoidance of possible pathogen carriers and potentially dangerous individuals (Abrams, Hogg, & Marques, 2005; Faulkner, Schaller, Park, & Duncan, 2004; Goodall, 1968; Kurzban & Leary, 2001; Neuberg & Cottrell, 2008; Park, Faulkner, & Schaller, 2003). Pathogen avoidance theory specifically proposes that physical anomalies are indicators of contagious disease, eliciting disgust feelings and avoidance behavior (Curtis, Aunger, & Rabie, 2004; Faulkner, Schaller, Park, & Duncan, 2004; Schaller & Duncan, 2007).

The universality of negative biases towards physical disabled or anomalous individuals indicates the presence of cognitive mechanisms necessary for the detection and evaluation of (body or face) anomaly or deviation. As the dislike of deviancy in simple patterns predicts dislike of individuals with physical disability (Gollwitzer et al., 2017), such a mechanism may be domain-general, and the individual differences in aversion may be transferred from one domain onto another.

Behaviour and social interaction

Behaviours violating social norms and expectations are judged negatively (Bond & Omar, 1992; Levine et al., 2000). Deviant or inappropriate behaviour can elicit negative reaction in onlookers (Leander, Chartrand, & Bargh, 2012; Szczurek, Monin, & Gross, 2012), and a lack of verbal-nonverbal consistency in communication reduces the likability of a first impression (Weisbuch, Ambady, Clarke, Achor, & Weele 2010). Thus, behaviours or interactions deviating from their familiar or predictable patterns are liked less. McAndrew and Koehnke (2016) suggest that behaviour is perceived as creepy if the intentions are ambiguous or potentially threatening. Furthermore, behaviour is perceived as creepy if it is deviating, especially when it violates social norms (Watt, & Maitland, & Gallagher; 2017). People with a higher aversion towards deviancy in simple patterns also exhibit a dislike of social norm

violations (Gollwitzer et al., 2022), again indicating domain-independent mechanisms and transfers between domains. Thus, ambiguous or deviating behaviour is associated with negative evaluations possibly due to a general dislike of deviation, anomaly, and ambiguity.

Negative judgments of social norm violations or deviant behaviour are common for people with mental conditions or neuroatypicality that impair the perception or execution of socially appropriate actions, like in autism-spectrum disorder (Alkhaldi, Sheppard, & Mitchell, 2019; Cunningham & Schreibman, 2009; Doherty-Sneddon, Whittle, & Riby, 2013; Lim, Young, & Brewer, 2022), attention deficit hyperactivity disorder (ADHD; Law, Sinclair, & Fraser, 2007; Nixon, 2001), or mental illness (Manago & Mize, 2022). Thus, negative evaluation of deviant behaviour may target vulnerable populations whose mental state does not allow for a proper following of social norms. The cause of such devaluations may lie in mechanisms described above, such as when intentions are ambiguous or unclear (McAndrew & Koehnke, 2016).

Creepiness in technology

A special branch of creepiness research is focussed on creepiness caused by new situations emerging from technological progress. Langer and König (2018) developed a scale to measure the creepiness of situations, defining creepiness as an unpleasant affective sensation elicited by unpredictable people and situations. Accordingly, unexplainable situations involving new technology are considered creepy (Langer, König, & Fitili, 2018; Langer, König, & Papathanasiou, 2019). Creepiness caused by technology is also investigated in the context of infiltrating a person's privacy (Tene & Polotensky, 2014). For example, technological recommendation systems (algorithms using personal data to recommend personalized articles) are perceived as creepy when their recommendations are accurate (Torkamaan, Barbu, & Ziegler, 2019). The perception of being tracked, observed, or assessed by technology can also cause creepy feelings (Pierce, 2019; Shklovski, Mainwaring, Skúladóttir, & Borgthorsson, 2014; Zhang & Xu, 2016). In summary, technology appears

creepy when it invades an individual's personal, intimate space, and when it is capable of collecting and using knowledge about an individual.

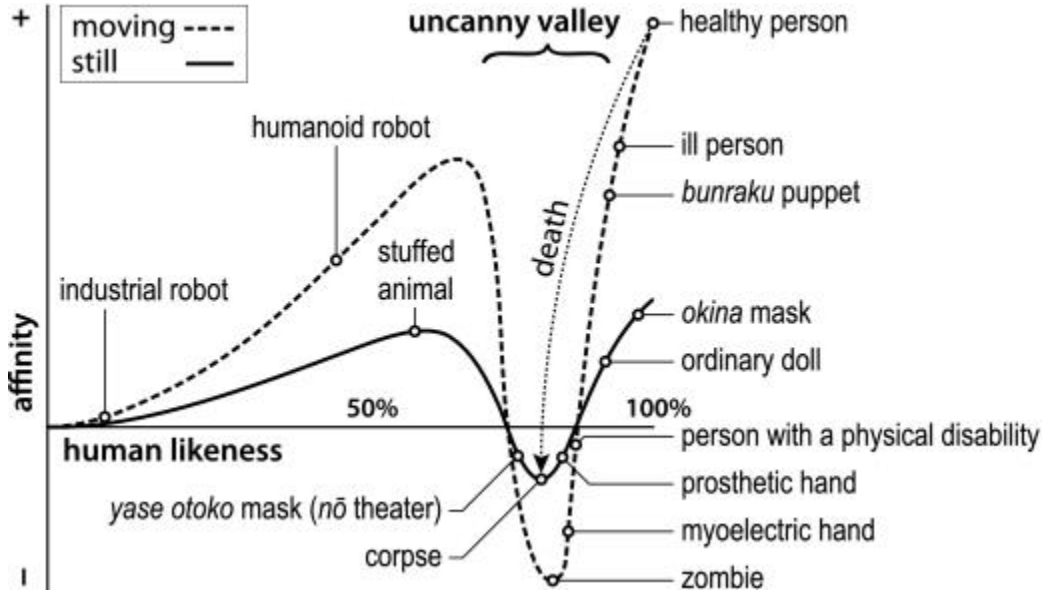
The uncanny valley

Accelerated technological advancement gave rise to increasingly humanlike artificial entities finding their place in human society. Social robots are used in various service roles such as healthcare, housekeeping, and waitering (Broekens, Heerink, & Rosendal, 2009; Dawe, Sutherland, Barco, & Broadbent, 2019; Lu et al., 2020; Nakanishi et al., 2020). Social robots can outperform human workers in situations that require performance beyond human abilities or when humans would be put at health risks: For example, social robots were implemented during the COVID-19 pandemic to minimize direct human-human contact (Aymerich-Franch & Ferrer, 2022). Similarly, computer-generated human imitations are used in a variety of settings, ranging from entertainment to advertisement, healthcare, and therapy (Salehi, Mehrabi, Fatehi, & Salehi, 2020). Recently, androids have been developed to replicate realistic human emotions (Sato et al., 2022). Thus, increasingly sophisticated hardware and developments in artificial intelligence are leading the way for highly realistic artificial humanlike entities to enter everyday human life and society.

However, some hurdles hinder the application of technology in everyday life: Entities at a certain threshold of near humanlike appearance are often seen as repulsive, strange, or eerie. Roboticist Masahiro Mori (1970) coined the term *uncanny valley* (UV) for a drop in likability for artificial entities in near levels of human likeness (compared to less or fully humanlike entities). According to Mori (1970), increasing the human likeness of an entity increases likability, but at a certain point at which likability drops into the negative (*Figure 1.1*).

Figure 1.1

The uncanny valley as proposed by Masahiro Mori (1970), translated by Karl MacDorman. Courtesy to Dr. Karl MacDorman.



Entities in this graphical "valley" are perceived as *uncanny*, *strange*, or *eerie*. At full human likeness, likability jumps back into the positive. Although Mori's (1970) first proposal was hypothetical, the UV function has been used as an explanation for various examples of people disliking humanlike androids or movie flops even leading to studio closures (Becker-Asano, Ogawa, Nishio, & Ishiguro, 2010; Freedman, 2012). The UV effect may also be a hindrance for trust-based interactions between humans and artificial agents (Mathur & Reichling, 2016). Specifically, the UV may reduce likability and trust in interactions with chatbots (Ciechanowski, Przegalinska, Magnuski, & Gloor, 2019), healthcare robots (Davies, 2016; Destephe et al., 2016; Olaronke, Rhoda, & Janet, 2017), and in video games (Tinwell, Grimshaw, Nabi, & Williams, 2011). Overcoming the UV effect is essential to avoid material risks through the inability to create trustworthy and likable realistic artificial entities, especially in a world where such entity gain increasing relevance.

In the following sections, the methodical and theoretical foundations on the UV effect will be reviewed. Two preliminary works by me and colleagues will be presented as a way to familiarize the reader with the main literature: For the methodological review, the meta-analysis by Diel, Weigelt, and MacDorman (2022) focussing on the variables used in previous UV research will serve as the foundation. For the theoretical review, the study by Diel and MacDorman (2021) investigating multiple theories will function as the major reference.

Research on the uncanny valley: A meta-analysis

A meta-analysis found that a UV effect can be consistently replicated with large effect sizes (Diel et al., 2022), although a relatively small number of studies failed to find a UV effect (e.g., Bartneck, Kanda, Ishiguro, & Hagita, 2009; Cheetham, Suter, & Jäncke, 2014; Cheetham, Wu, Pauli, & Jäncke, 2015). The uncanny valley is typically investigated using two variables, each representing an axis of Mori‘s (1970) curve: human likeness as the independent variable and an aesthetic measure (usually likability, eeriness, or a similar measure) as the dependent variable.

Human likeness

For the independent variable, human likeness or a related variable like humanness, realism, anthropomorphism, or – if animal stimuli are used – zoomorphism is typically used (Diel et al., 2022). Human likeness is manipulated in different ways depending on the study, although it is usually explicitly measured as a self-assessment rating scale or index (Diel et al., 2022). While a combination of scales or a questionnaire can more accurately assess the construct of human likeness, previous research can reliably replicate a UV effect using single scale items only (Diel et al., 2022).

Despite the initial focus on human likeness in the UV model (Mori, 2012), it is important to note that the UV effect can be reliably found using animal stimuli (Diel & MacDorman, 2021; Löffler, Dörenbächer, & Hassenzahl 2020; Schwind et al., 2018; Yamada et al., 2013). Hence, the human-specificity of the UV is questionable, and the effect may be generalizable onto other categories. In that case, then the variable *human likeness* would be insufficient to capture the generality of the phenomenon. Whether replication of UV curves in non-humanoid entities should be labelled as an *uncanny valley proper* can be a semantic or conceptual debate – arguably, the UV effect has been used for humanlike entities specifically and should be used for those only, even if analogous statistical patterns can be observed in other categories. On the other hand, if the UV effect can be reconceptualized as a more general phenomenon, then such a model would be capable of explaining a wider range of data despite similar complexity and thus should be preferred from a scientific standpoint. The generalizability of the uncanny valley beyond human likeness (or animal likeness) remains an open question.

According to Diel et al. (2022), stimuli varying on human likeness (or analogous concepts) can be categorized into different creation techniques, with relevant ones described now. The most common type of stimulus set, *distinct entities* is a collection of stimuli representing different (e.g., mechanical robot, CG characters, android, human) entities, often taken from image search websites (e.g., Mathur et al., 2020). While this approach uses real-world stimuli that tend to be ecologically valid for the uncanny valley (e.g., androids), there is often little control between the stimuli, leading to potential confounding variables like differences in lightning, colour, or emotional expressions (Diel et al., 2022). Another technique, *morphing*, uses faces of two distinct entities (usually a human and a robot or CG face) and creates gradual transitions from one face to another, leading to a range of stimuli differing on comparable levels of change (Diel et al., 2022). Morphing however may lead to morphing

noise and stimuli that would not occur in the real-world (e.g., MacDorman & Ishiguro, 2006; Seyama & Nagayama, 2007). Similarly, *face distortion* techniques incrementally change features or configurations in faces, creating a range of stimuli with incremental distortions (e.g., Green et al., 2008; MacDorman et al., 2009; Mäkäräinen et al., 2014). Similar to morphing, stimuli created with this technique would not naturally occur and may not represent stimuli relevant for the uncanny valley in e.g., androids or CG characters. *Realism render* techniques create ranges of stimuli by using the same entity or actor and manipulating the level of depicted realism, for example through rendering the image or video through editing filters to remove detail (e.g., MacDonnell et al., 2012). Finally, *real-life encounter* is the arguably most ecological valid as it introduces participants to actual androids, often to simulate social interaction (e.g., Zlotowski et al., 2015). Some techniques, like distortion, have been used on other categories or modalities: For example, voices have been distorted to create a range of voices differing on a level of human likeness (e.g., Baird et al., 2018). Thus, the stimulus creation techniques described by Diel et al. (2022) can be used for different stimulus categories and processing domains.

Uncanniness

Mori's (2012) original terms to describe the UV effect, *shinwakan* and *bukimi*, have been translated as affinity, familiarity, warmth, or likability on the one hand and eeriness or uncanniness on the other (Bartneck et al., 2009; Diel et al., 2022; Ho & MacDorman, 2010). Emphasis has been put on uncanniness or eeriness as a specific negative sensation characteristic to the UV effect (Diel et al., 2022; MacDorman & Entezari, 2015; Mangan, 2015). The specific sensation has been associated with fear and disgust (Ho, Pradomo, & MacDorman, 2008), and described as a form of anxiety caused by an inability to conceptualize information (Mangan, 2015). Recently, Benjamin and Heine (2022) developed

an index to measure an *uncanny feeling* presumed to be caused by a variety of circumstances including but not limited to the uncanny valley.

Given the lack of uniform translation, it is not surprising to see the heterogeneity of measurements used for the dependent variable in UV research: methods range from self-report scales to behavioural measures like avoidance (Diel et al., 2022). Within self-report measures, measures with the highest effect sizes when UV effecrs are observed are *threatening*, *likable*, *aesthetic*, *familiar*, and *eerie* (Diel et al., 2022). However, specific negative experience like *creepy*, *eerie*, or *uncanny* would be more precise in measuring the UV effect by being devoid of confounding factors that may affect more general items (e.g., *likable*), while still replicating large effect sizes (Diel et al., 2022). Such affect measures may also be beneficial over behavioural measures as the latter may not necessarily capture the "uncanny experience"; for example, participants may not want to interact with a robot not because it is uncanny but because it is boring; alternatively, different trajectories for affect ratings and behavioural measures are occasionally found (e.g., Strait et al., 2015). Thus, negative affect measures on specific experiences like *eerie* or *uncanny* are best suited to measure the distinct feeling associated with the proper uncanny UV effect.

Statistical representation of the uncanny valley

Statistically, the UV effect is usually investigated either by group comparisons between groups presumably or verifiably varying on a human likeness axis (e.g., mechanical robots, androids, and humans), or by plotting polynomial curves resembling a U- or N-shaped function akin to Mori's (2012) curve (Diel et al., 2022). The exact statistical method depends on the research design: for example, using clear categories of stimuli (e.g., a group of robot stimuli and a group of human stimuli), or a very limited range of stimuli allows for group or stimulus comparisons with the prediction that stimuli or categories on the intermediate human likeness level are more uncanny or less likable than the other stimuli (e.g., Diel &

MacDorman, 2021). If the stimulus range is large and not clearly categorizable, polynomial regressions or similar methods are favoured (e.g., Mathur et al., 2020). Both methods are acceptable in UV research (Diel et al., 2022).

In summary, UV research typically focusses on one independent variable (typically human likeness) and one dependent variable (usually an affect variable), which are measured with self-report or behavioural measures in most studies. Stimuli varying on human likeness are created through a variety of methods, like selecting images of distinct entities, morphing, rendering, distortion, or interactions with real-life robots. A large variety of the dependent variables exist, with theoretical and empirical arguments favouring items measuring specific negative sensations related to the "uncanny feeling", like *eerie*, *uncanny*, *creepy*, *strange* or *weird*. Behavioural measures tend to be unspecific and should be used alongside self-report items. Depending on the methods used, group comparisons or polynomial regressions are viable to test for the UV effect.

While this section has focussed on how to investigate the UV effect as a working hypothesis, the following chapter will investigate theories on the effect.

Theories on the uncanny valley

Since Mori's (2012) initial proposal of the uncanny valley in 1970, A wide range of theories on the UV effect have been proposed and investigated. Multiple papers have already focussed on the different explanations, and an exhaustive summary and ordering of the theories will be presented here that is mostly based on previous summaries (Diel & MacDorman, 2021; Diel et al., 2022; Kätsyri et al., 2015; MacDorman et al., 2009; Wang, Lilienfeld, & Rochat, 2015). The theories will be ordered based on their explanatory level (feature-level, cognitive, affective-motivational, evolutionary, and neural).

Feature-based theories

Feature-based theories predict the that the uncanny valley occurs due to the presence of specific features within a stimulus. These theories make no assumptions to underlying cognitive or neural mechanisms, although can be linked to those. All feature-based theories are domain-general and thus not exclusive to human(-like) stimuli.

Atypicality. Atypicality theories predict that any kind of anomalous, deviant, or distorted features elicit uncanniness (Kätsyri et al., 2015). *Atypicality* is similar to, albeit more general than other feature-based theories.

Mismatch. Mismatch theories predict that entities containing mismatching features (i.e., features taken from different stimulus groups or categories) elicit uncanniness (Seyama & Nagayama, 2007). This includes multimodal mismatch, such as a mismatch between voice and appearance (Mitchell et al., 2011).

Realism inconsistency. Realism inconsistency theory can be seen as a more specific variation of mismatch, as it predicts uncanniness caused by entities containing features with different levels of realism (MacDorman & Chattopadhyay, 2016).

Specialized (e.g., configural) processing. The final theory refines previous feature-level theories by attuning for a general weakness among them: the moderating effect of stimulus category. For example, face distortions appear more uncanny when faces are more realistic (Mäkäräinen et al., 2014; MacDorman et al., 2009), or in human faces compared to cat faces or houses (Diel & MacDorman, 2021). However, the previous feature-based theories do not attempt predictions or explanations on why the sensitivity to deviations is higher in some categories. The degree of specialized processing used for a certain type of stimulus may moderate this sensitivity to distortions, so that feature-level distortions may be more apparent and more uncanny in stimulus categories exhibiting a higher level of specialization and thus

more detail-based processing (Diel & MacDorman, 2021). Realistic human faces are such a category recruiting highly specialized processing (Farah, Tanaka, & Draom, 1995; Rhodes, Brake, Taylor, & Tan, 1989; Richler, Cheung, & Gauthier, 2011). Thus, this theory extends feature-based theories by Adding a moderator variable: a higher degree of specialization should cause increased sensitivity to distortions compared to the base stimulus.As the dissertation will focus on this theory, specialized processing and its relation to the UV effect will be discussed in more detail further below.

Cognitive theories

Cognitive explanations of the UV rely on pre-established theories of human cognition. Although cognitive theories are based on established theories, a common issue is the inability to explain the specific sensation of *eerie* or *uncanny* which is characteristic for the UV effect, and instead, if at all, only explain general decreases of likability (Diel & MacDorman, 2021; MacDorman & Entezari, 2015; Mangan, 2015).

Categorization-based theories. Multiple theories explained the UV effect using categorization-based processes. Categorically ambiguous tend to be disliked, especially when participant have to attend to the ambiguous category dimension (e.g., gender in a gender-ambiguous face; Halberstadt & Winkielman, 2014; Owen et al., 2016; Winkielman et al., 2015). It has also been proposed that categorically ambiguous stimuli, i.e., those difficult to categorize, elicit uncanniness (Cheetham et al., 2013). Uncanny stimuli can be difficult to categorize (Ferrey et al., 2015; Kawabe et al., 2017; Yamada, Kawabe, & Ihaya, 2013), and when participants are asked to attend to the ambiguous category of human likeness when rating androids, androids are rated as less likable (Carr et al., 2017). Thus, not only were correlations between categorical difficulty and uncanniness found, but focusing on the categorical ambiguity also increases uncanniness of androids.

However, the most uncanny stimuli may not necessarily be the most categorically ambiguous and vice versa (MacDorman & Chattopadhyay, 2016; Mathur et al., 2020). However, while the humanoid stimuli used by MacDorman and Chattopadhyay (2016) and Mathur et al. (2020) have been categorized on based their humanness, they may have been ambiguous on other dimensions (e.g., facial expression) which could have caused uncanniness instead.

In sum, research findings on the association between categorical difficulty and uncanniness are inconsistent and difficult to interpret, and the link between those variables remain unclear.

Cognitive dissonance. Cognitive dissonance is an aversive state caused by entertaining contradicting cognitions (Festinger, 1957). Androids may elicit contradicting beliefs, such as classifications as both "human" and "robot", or "living" and "inanimate" (Hanson, 2005; MacDorman & Entezari, 2015; Tondu & Bardou, 2011). As contrasting cognitions may be competing categories, this theory is compatible with categorization-based explanations. Higher ratings of human uniqueness correlate with uncanniness of androids, which may be due to increased dissonance created by the stronger belief in human uniqueness (MacDorman & Entezari, 2015). However, experiments manipulating cognitive dissonance to investigate the UV effect have not yet been conducted.

Cognitive load. Similar to the previous theories, it has been suggested that uncanny stimuli increase cognitive load, for example by activating cognitive conflict between competing categories (Weis & Wiese, 2017; Yamada et al., 2013). Again, cognitive load is compatible with categorization-based theories, but can be extended to instances of uncanny stimuli which are not categorically ambiguous but increase cognitive load for another reason.

Dehumanization. Initially proposed by Wang, Lilienfeld, and Rochat (2015), dehumanization theory proposes that the UV effect is caused by first recognizing an artificial humanlike entity as human, then dehumanizing it by removing humanlike attributions based on its artificial

features. Attribution of animacy decreased 400ms after android stimuli were presented, supporting the notion of dehumanization (Wang, Cheong, Dilks, & Rochat, 2020). In addition, disrupting the proposed process by pre-emptively dehumanizing androids reduces the UV effect (Yam, Bigman, & Gray, 2021). However, as the dehumanization hypothesis is primarily focused on recognizing the stimuli as humanlike in the first place, the hypothesis would have difficulties predicting uncanniness in stimuli varying on a dimension other than human likeness, such as animal stimuli or inanimate objects. Thus, dehumanization is a category-dependent explanation focusing on a human likeness variable specifically.

Exposure frequency. Some researchers have criticized the idea that the UV effect may be due to a specific cognitive mechanism. Instead, Burleigh and Schoenherr (2015) suggested that the effect is caused by a relative lack of exposure to stimuli on intermediate levels of human likeness (e.g., morphed faces or androids) compared to more robotic or human faces. As exposure is known to increase likability (Zajonc, 1968), relatively low-exposure stimuli would also be less likable.

Inhibitory devaluation. Ferrey, Burleigh, and Fenske (2015) proposed that the UV effect is caused by cognitive inhibition elicited to solve a cognitive conflict of a stimulus triggering competing representations.

Misattribution. Gray and Wegner (2012) proposed that uncanniness is caused by attributing humanlike qualities like mind onto clearly nonhuman entities (e.g., a supercomputer). Although this explanation has gained some support (e.g., Appel, Izydorczyk, Weber, Mara, & Lischetzke, 2020; Müller, Gao, Nijssen, & Damen, 2020; Stein & Ohler, 2017), other studies found evidence against misattribution of human qualities (e.g., Wang et al., 2020). Again, the misattribution hypothesis is specific to human(-like) stimuli and cannot easily predict a UV effect in non-human and especially inanimate categories.

Processing disfluency. Processing fluency theory predicts that prototypical stimuli are easily processed and thus appealing (Halberstadt & Winkielman, 2013; Oppenheimer, 2008; Winkielman et al., 2003). Ambiguous stimuli however lead to processing disfluency, which elicits negative affect (Halberstadt & Winkielman, 2014). Similarly, stimuli deviating from the typical appearance (thus, stimuli unlikely to statistically occur) have also linked to increased processing cost (Dotsch et al., 2016; Ryali et al., 2020; Vogel et al., 2020). Thus, processing disfluency is compatible with both category-based explanations of the UV effect as well as feature-based accounts (atypicality, mismatch, specialized processing) as atypical, deviating, or mismatching features would be a cause of processing disfluency. The effect of perceptual disfluency depends on the expectations of typical appearance which often depend on previous experience (Wänke & Hansen, 2015). Thus, processing disfluency is also compatible with the view that perceptual specialization with a stimulus category may also increase the sensitivity to deviations caused by a higher degree of disfluency.

Affective-motivational

Affective-motivational theories on the uncanny valley include a wide range of explanations focussing on specific negative states and behaviours caused by processes including, but not limited to cognitive processes.

Mortality salience. Terror management theory predicts that becoming aware of one's mortality (mortality salience) activates unconscious defence mechanisms to reduce anxiety and promote self-preservation (Greenberg, Pyszynski, & Solomon, 1986; Pyszczynski, Solomon, & Greeenberg, 2015). Hence, it has been suggested that death-indicating features of uncanny androids (e.g., pale skin, lifeless eyes) may activate mortality salience, leading to a dislike of the android (Koschate, Potter, Bremner, & Levine, 2016; MacDorman, 2005). MacDorman (2005) found that viewing uncanny robots increased preference for people who

support one's worldview, as predicted by terror management theory. In addition, Koschate et al. (2016) found that uncanny stimuli increased the accessibility to death-related thoughts.

However, the relation between uncanniness and mortality salience remains unclear: Are androids uncanny because they elicit death-related thoughts, or is uncanniness caused by another mechanism and then elicits death-related thoughts by increasing stress or anxiety? Further research is needed to investigate the exact causal relationship.

Novelty avoidance. Sasaki, Ihaya, and Yamada (2017) proposed that uncanny stimuli are disliked because they are novel due to being hard to categorize, and found that behavioural inhibition, a trait associated with increased anxiety towards novel stimuli, predicts the severity of uncanniness ratings.

Psychopathy avoidance. Proposed by Tinwell and colleagues (2013), psychopathy avoidance predicts that indicators of social or emotional inauthenticity in an android or CG face elicit uncanniness due to defence mechanisms for the detection of psychopathic, deceptive, or malevolent intent. This theory is domain-specific as psychopathy avoidance mechanisms can hardly be generalized onto non-human categories.

Threat to human identity. Artificial humanoid entities may be a threat to how people define humans as a unique species. Kaplan (2004) suggested that increasingly humanoid robots trigger the need to redefine what it means to be human. Belief in human uniqueness is furthermore associated with uncanniness ratings of androids (MacDorman & Entezari, 2015; Ratajczyk, Dakowski, & Lupkowski, 2023), and Ferrari, Paladino, & Jetten (2016) showed that uncanny androids are also judged as threatening to the human-robot distinction. However, the exact causal relationship is not yet understood. In addition, it is not clear whether it is the threat to specifically *human* uniqueness that may cause uncanniness, or whether it is more, generally, categorical ambiguity or contradicting categorizations. As

threat to human identity is an explanation specifically to stimuli varying on a human likeness axis, it would have difficulties explaining an UV effect in other stimulus categories like animals or objects.

Evolutionary

Evolutionary theories provide explanations on the distal causes of the UV effect and why humans may have evolved a UV reaction as an adaptive strategy for survival and procreation. Evolutionary theories on the uncanny valley however have been criticized for escaping empirical falsification (Urgen et al., 2018).

Disease avoidance. Mori (1970) initially suggested that the uncanniness of humanlike entities may stem from their similarities to dead or diseased human bodies. Accordingly, *disease avoidance* theory explains the UV effect as an evolved mechanism to detect and avoid indicators of contagious disease (MacDorman & Ishiguro, 2006). Disgust can be understood as an evolutionarily beneficial response to prepare the body and mind to avoid contamination (Chapman & Anderson, 2012; Rozin & Fallon, 1987). Fittingly, disgust and disgust sensitivity have been associated with the uncanniness of humanlike entities (Ho, MacDorman, & Pramono, 2008; MacDorman & Entezari, 2015). MacDorman and Entezari (2015) have furthermore suggested that the disgust component of uncanniness may be related to its specific subjective experience. However, this explanation can only be applied to animate categories (humans and animals) and not to uncanniness caused by objects.

Mate selection. MacDorman and Ishiguro (2006) proposed that uncanny androids may carry indicators of low fertility or bad genetic fitness. As indicators of fitness are linked to face attractiveness (Rhodes & Zebrowitz, 2002), uncanniness may be an evolutionary response triggered by features indicating an unfit mating partner. Again, this theory is restricted to human or arguably animal stimuli (see also Diel & MacDorman, 2021).

Neural correlates

Finally, neural theories focus on neurocognitive mechanisms that may underlie the UV effect.

Expectation violation/prediction error. The predictive coding framework envisions the brain in an efficient equilibrium when internal generative models and predictions of the world correspond with the sensory input, while a discrepancy between prediction and input elicits a prediction error (Friston, 2010; Keller & Flogel, 2018). Costs of prediction errors can be minimized by avoiding potentially surprising events and by updating the internal representation of the world to fit the new input (Friston, 2010). Prediction errors are typically operationalized as increased neural activity observed when sensory input conflicts expectations (Fiser et al., 2016; Makino & Komiyama, 2015; Meyer & Olson, 2011).

The N400 event-related potential (ERP) component has been established as a neural correlate of predicitons error (Kutas & Federmeyer, 2011). For the most part, increased N400 components are observed for unexpected events or semantic errors in written words or sentences (Kutas & Hillyard, 1980). N400 effects have also been observed for face stimuli, for examples mismatches between familiar faces and learnt context primes (Jemel, George, Olivares, Fiori, & Renault, 1999; Olivares & Iglesias, 2010; Wiese & Schweinberger, 2008). N400 effects correlated with incongruent facial structure and identity are rooted in face-sensitive functional areas and thus may reflect face processing mechanisms (Olivares et al., 2018). In the context of the UV effect, prediction errors could occur when experience-driven expectations of human appearance and behaviour disagree with the observation of an artificial android. Indeed, discrepancies between androids' human appearance and mechanical motion elicit activity associated with prediction errors, such an increased N400 components (Mustafa, Guthe, Tauscher, Goesele, & Magnor, 2017; Saygin et al., 2012; Urgen et al., 2018). However, previous research on N400 components and the UV effect did not directly measure the stimuli's uncanniness, thus the link between the component and

uncanniness itself remains unclear. In fact, Urgen et al. (2018) only found android N400 effects for moving stimuli, but no differences between N400 amplitudes between still images of androids, robots, and humans, despite it being known that even still images of androids are uncanny. Thus, N00 amplitudes as indicators of prediction errors seem to not fully explain the UV effect, or at least uncanniness only in relation to a discrepancy between motion and appearance. In fact, no study has yet linked indicators of prediction errors to the uncanniness of still images.

Neural correlates of the uncanny valley. Rosnethal-von der Pütten, Krämer, Maderwald, Brand, & Grabenhorst (2019) found a function of human likeness in neural activity in the ventromedial prefrontal cortex (VMPFC) analogous to Mori's (1970) uncanny valley curve when participants were shown images of robots, androids, and humans. The authors argued that the VMPFC signal was computed from two separate sources, specifically the temporoparietal junction coding linearly for human likeness, and the fusiform gyrus coding for human vs non-human distinction.

Deviation and familiarity: A refined theory of the uncanny

Falsification is the driving force of scientific progress, yet research on the UV effect suffers from an inflation of explanations and theories. The goal of this dissertation is thus twofold: First, it is to critically evaluate the variety of prevalent theories on the UV effect. Second, it is to conceptualize a refined and encompassing theory of the UV effect that explains previous findings.

Traditionally, the UV effect is conceptualized as a polynomial, *N*-shaped function between human likeness and uncanniness (*Figure 1A*). However, polynomial functions beyond quadratic ones are rare in nature and psychology specifically, and may instead be caused by moderating influences of underlying variables. Such alternative interpretations of the data

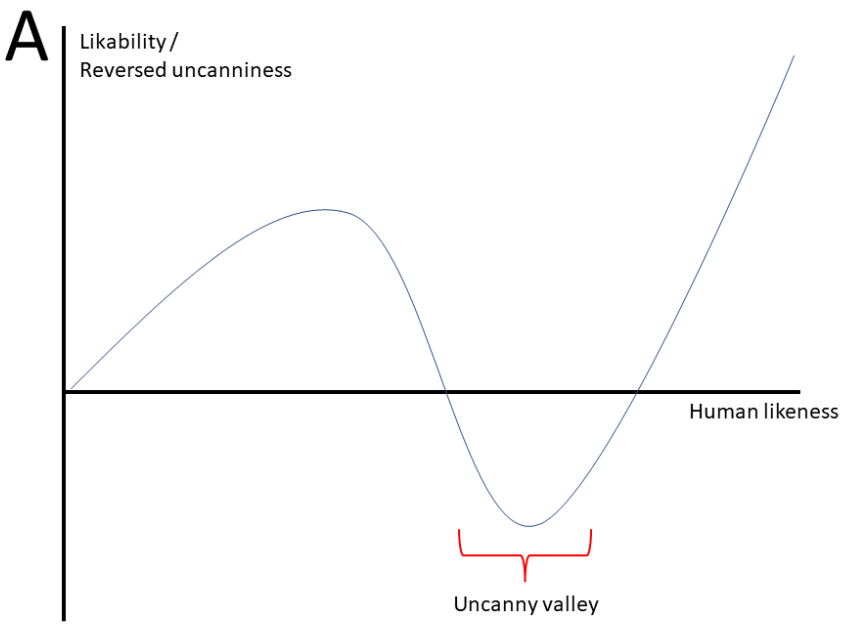
have been implicitly indicated by previous researchers: For example, the UV effect itself has been often described as an increased sensitivity to errors or nonhuman features that increases with the entity's realism or human likeness (Chattopadhyay & MacDorman, 2016; Green et al., 2008; MacDorman et al., 2009). In this sense, a decrease of likability (or increase of uncanniness) would be caused by deviations in a linear manner, and this effect would be stronger on more realistic or humanlike entities, creating a moderated linear function of likability, level of deviation, and realism or human likeness. Indeed, the facial distortions appear more uncanny in more realistic faces (MacDorman et al., 2009; Mäkäräinen et al., 2014), and in human faces compared to cat faces or houses (Diel & MacDoramn, 2021), indicating the role of a third moderating variable related to the stimulus' category. Unrealistic proportions may be acceptable for a mechanical humanoid robot, but with an increasing level human likeness, even small deviations in an android's face may appear uncanny. Thus, errors and deviations, which may occur during the design of artificial humanoids, are easily detectable in more humanlike artificial entities.

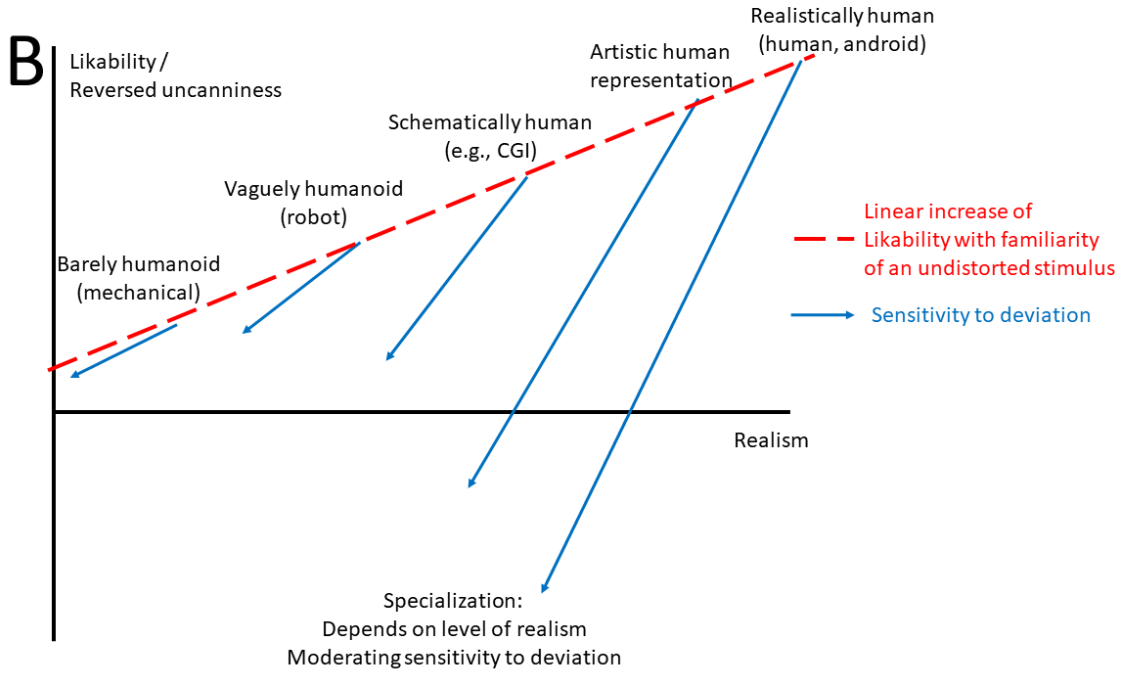
How would an increase of human likeness and realism increase the sensitivity to deviations? Specialized processing develops through prolonged exposure to a type of stimulus and increases the ability to differentiate individual exemplars of a stimulus that may otherwise be difficult to differentiate (Lee, Anzures, Quinn, Pascalis, & Slater, 2011; Rhodes, Brake, Taylor, & Tan, 1989). One way to measure specialized processing in faces is the *inversion effect*: As faces are usually experienced upright, humans show a higher degree of recognition performance for upright compared to inverted faces (Carbon & Leder, 2006; Kanwisher & Moscovitch, 2000; Maurer, Le Grand, & Mondloch, 2002). CG faces meanwhile show a lower inversion effect than realistic faces (Crookes et al., 2015), indicating a lower degree of specialized processing. As specialized processing eases the ability to detect differences between stimuli, it may also enhance the ability to detect deviations that may then be

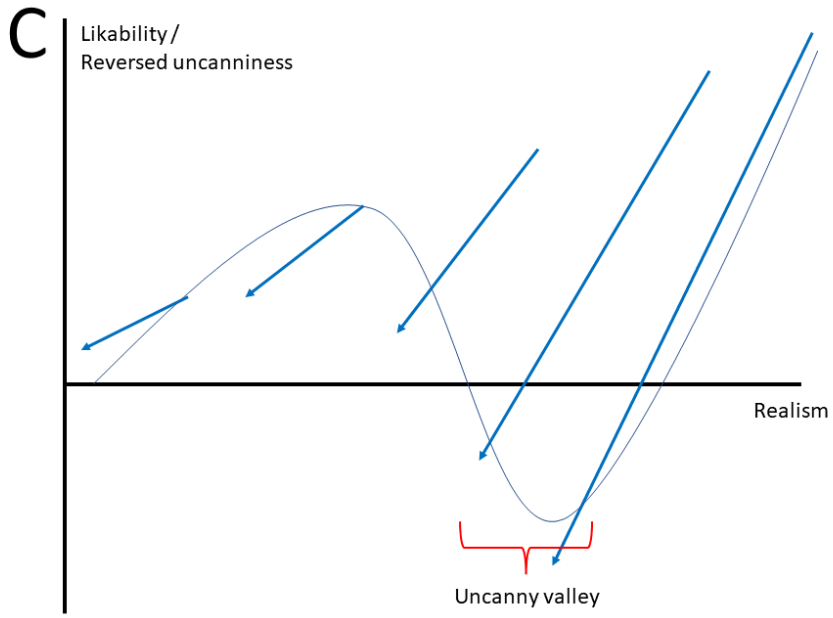
aesthetically devalued. Indeed, face inversion reduces the variation of aesthetic evaluation (Bäuml, 1994; Leder, Goller, Forster, Schlageter, & Paul, 2017; Santos & Young, 2008). Thus, a moderated linear function of likability, deviation, and specialized processing may hypothetically appear as presented in *Figure 1.2B.* According to this view, realistic human faces would activate the highest level of specialized processing which in turn would sensitize strongest for deviations. Given their similarity to real human faces, android faces would thus be judged on a manner similar to human faces, pushing them into an uncanny valley. Less realistic faces meanwhile would enjoy a wider range of acceptable variation due to a decreased level of specialization and error sensitivity. Because specialized processing would correlate with human likeness or realism, the data of a moderated linear function should also be able to be plotted as a polynomial function akin to the uncanny valley (*Figure 1.2C*).

Figure 1.2

Different functions to describe data varying across dimensions of human likeness and likability or uncanniness. Figure 1.2A shows a prototypical uncanny valley plot as established by Mori (1970). Figure 1.2B depicts an alternative interpretation of the data: A higher level of realism of a base stimulus elicits a higher level of specialized processing, increasing the sensitivity to (and thus uncanniness of) deviations. Figure 1.2C shows how the same data may be plotted as either a polynomial function or a moderated linear function.


A
Likability /
Reversed uncanniness
Human likeness
Uncanny valley


B
Likability /
Reversed uncanniness
Realistically human
(human, android)
Artistic human
representation
Schematically human
(e.g., CGI)
Vaguely humanoid
(robot)
Barely humanoid
(mechanical)
Linear increase of
Likability with familiarity
of an undistorted stimulus
Sensitivity to deviation
Realism
Specialization:
Depends on level of realism
Moderating sensitivity to deviation


C
Likability /
Reversed uncanniness
Realism
Uncanny valley

While the refined theory is a conceptual prediction of the effect of specialization on the sensitivity to deviation, it can also be expressed mathematically. Specifically, the moderated linear function can be expressed as a moderation function of uncanniness (U) of a deviating stimulus (d_x), moderated by the degree of specialization for the stimulus or stimulus category (s_x):

$$U(d_x) = s_x d_x$$

With $U(d_x)$ as the uncanniness of the deviating stimulus d_x, with d referring to the level of physical deviation of the stimulus x relative to the typical variation of stimulus x (0_x would refer to a non-deviating stimulus or a stimulus falling within the typical variation of stimulus type x, and $U(0_x)$ to the uncanniness of a non-deviating stimulus). The variable s_x refers to the degree of specialization towards stimulus x. Thus, increasing deviation of stimulus x would increase uncanniness, which is furthermore increased by the level of specialization for stimulus x.

Because the total value of uncanniness would also depend on the uncanniness of a non-distorted stimulus (e.g., a non-distorted realistic human face is expected to be less uncanny than a non-distorted robot face), a more accurate mathematical expression of the formula includes such a term:

$$U(d_x) = U(0_x) + s_x d_x$$

If x represents the stimulus type *realistic human face* (see *Figure 1.2B*), android faces would be considered deviating examples of this stimulus category (d_x). Because of the high level of specialization towards realistic human faces (represented by a higher s_x value), android faces would be particularly uncanny according to this model compared to the uncanniness of non-deviating realistic human faces (0_x). Deviations in stimulus categories with a lower degree of specialization (s_x value), like mechanical robots, would meanwhile not increase uncanniness

as much compared to the non-deviating stimulus. Such a mathematical expression would provide additional predictions, such as higher beta values of the effect of deviation on uncanniness for more specialized categories.

One way to investigate this *moderated linear function hypothesis* is to test whether a set of data that can be plotted as an uncanny valley would be better explained by a linear moderated function as described above. This can be done by using indicators of specialization and deviation for different stimulus types and testing a linear moderated model (see Chapter 3). Indirectly, this model can be investigated by investigating whether the difference in uncanniness between a deviating and non-deviating stimulus is increased for more specialized categories (e.g., Chapters 2 and 4).

However, because variables like the degree of deviation have not been measured or manipulated in previous UV research, new research would need to create sets of stimuli varying on levels of realism and deviation.

Humans have a natural aversion to deviancy even in simple patterns (Gollwitzer et al., 2017). If uncanniness is caused by deviations, then deviancy aversion as a individual-level variable should predict the intensity of the UV effect. Specialization may recruit additional dimensions of information (e.g., configural information faces), which increases the likelihood to detect deviations that may be negatively evaluated. A cumulation of aversion caused by deviation on multiple dimensions would then lead to a drastic decrease of likability, or an increase of uncanniness. Such an explanation would be in tune with cognitive or neural theories like processing disfluency (Winkielman et al., 2003) and expectation violation (Friston, 2010). The negative evaluation of naturally occurring deviations of human appearance, such as those observed in physical disabilities, could also be explained by such a theory.

Thus, the UV effect can be rethought of as a moderated linear function between familiarity or specialization, deviation, and likability, which can also be plotted as a polynomial function when plotted against human likeness or realism. However, as familiarity or specialization is more domain-general than human likeness, this refined uncanniness theory could also be applied to non-human stimuli like written text or physical places (see Diel & MacDorman, 2021), which will be explored now.

Deviation and aversion across different domains.

Body processing. Evidence for specialized processing (e.g., inversion effect for recognition) has been observed for bodies (Keye, Mingming, Tiantian, Wenbo & Weiqi, 2019; Reed, Stone, Bozova, & Tanaka, 2003; Stekelenburg & de Gelder, 2004). In addition, specific neural substrate have been correlated with body processing, specifically the fusiform and extrastriate body areas (FBA and EBA; Astafiev, Stanley, Shulman, & Corbetta; Peelen & Downing, 2005; Taylor et al., 2009; van der Riet, Grèzes, & de Gelder, 2009). The FBA especially has been associated with the configural processing of bodies (Brandman & Yovel, 2016). Effects of body deviations on aesthetic ratings or neural activity however is, as of yet, lacking.

Voice processing. Mechanisms underlying the processing of deviating faces may be transferable onto voices, as face and voice processing share many similarities (Belin et al., 2011; Young, Frühholz, & Schweinberger, 2020; Schweinberger, Kawahara, Simpson, Skuk, & Zäske, 2014). Similar to faces, individual voices vary on structural dimensions (e.g., *formant frequencies*: frequency peaks resulting from resonances in the vocal tract). Such dimensions are used to differentiate individual voices (Baumann & Belin, 2008; Gaudrain, Li, Ban, & Patterson, 2009; Latinus et al., 2013; Lavner, Gath, & Rosenhouse, 2000; Schweinberger et al., 2014). Deviating voices are perceived as more distinct and elicit stronger BOLD signals in voice-specific neural substrates (Andics et al., 2010; Latinus et al.,

2013), which may indicate increased processing need for atypical stimuli. Such deviating voices may also suffer from negative evaluation of aesthetic appeal: Disorders affecting voice (e.g., Reinke's edema, muscle tension dysphonia) increase the perception of voice atypicality (Kreiman, Auszmann, & Gerrat, 2018; Kreiman & Gerratt, 2003; Kreiman, Gerratt, Precoda, & Berke, 1992) and are evaluated more negatively across various social dimensions compared to healthy voices (Altenberg & Ferrand, 2006; Amir & Lavino-Yundof, 2013; Schroeder, Rembrandt, May, & Freeman, 2020). Thus, deviating voices may also fall into an uncanny valley of voices.

Written word processing. Analogous arguments can be made for the processing of written words. Words written in a familiar language are processed holistically (Pelli, Farell, & Moore, 2003), analogously to faces (Martelli, Maja, & Pelli, 2005). The neural substrate contralateral to the face-sensitive FFA has been associated with the processing of words and letter strings, called the visual word form area (VWFA; Dehaene & Cohen, 2011; Dien, 2009; Hillis et al., 2005). Evidence suggests configural processing of words (Barnhart & Goldinger, 2013; Björnström, Hills, Hanif, & Barton, 2014; Gauthier & Wong, 2006; Wong, Twedt, Sheinberg, & Gauthier, 2010), and its disruption in dyslexia (Conway, Brady, & Misra, 2017). Wong et al. (2019) found that participants are sensitive to subtle changes in a word's configuration (e.g., slightly misaligning Latin letters or radicals of a Chinese character),but only when participants were familiar with the language and when words were presented upright instead of inverted. Thus, observers are sensitive to even subtle changes of a familiar language's configural pattern but only if the configuration is intact. It is an open question however whether configural deviations within a word would also elicit uncanniness, as expected from the hypothesis that deviations in specialized categories were to cause uncanniness.

Place processing. Multiple neural functional areas have been identified for place processing, most notably the parahippocampal place area (PPA; Epstein & Baker, 2019; Epstein & Kanwisher, 1997; Stansbury et al., 2013). Increased PPA activity has been found for contextually incongruent scenes compared to congruent ones (Rémy et al., 2013).

Some physical places can elicit feelings of creepiness (McAndrew, 2020). A high degree of "mystery" may be elicited by places not allowing inference of sufficient information, for example in environments with fog or dim light (Kaplan, 1987; Stamps, 2007). Schema-based typicality may be another source of information (Widmayer, 2002). Typical physical places follow predictable configural patterns (e.g., positions and number of doors, windows, furniture), and places deviating from these patterns may appear uncanny. Inconsistent scenes are less likable (Shir, Abudarham, & Mudrik 2021), just as built environments lacking coherence (Coburn et al., 2020; Cartanian, Navarrete, Palumbo, & Chatterjee, 2021; Weinberger, Christensen, Coburn, & Chatterjee, 2021). Finally, houses with distorted features have been shown to elicit uncanniness (Diel & MacDorman, 2021). In summary, it can be argued that physical places follow predictable configural structures and that deviations from these structure decrease appeal and neural processing need.

Dynamic face emotion expression processing. Humans can infer the emotional state of another by analyzing even subtle changes in facial expressions (Bould & Morris, 2008). As facial expressions are marked by changes in configural information, more specifically the motion of face action units (AUs; Ekman & Friesen, 1978), emotional facial expressions may be processed configurally (Martinez, 2017). Multiple studies found that disrupting the configuration of an emotionally expressive face (e.g., through inversion) reduces the accuracy to recognize the expressions (Bartlett & Searcy, 1993; Bombari et al., 2007; Calder et al., 2000; Calder & Jansesn, 2005; Calvo & Nummenmaa, 2008; Derntl et al., 2009; Durand et al., 2007; Fallshore & Bartholow, 2003; Goren & Wilson, 2006; McKelvie et al., 1995;

Pollak et al., 2009; Searcy & Bartlett, 1996). Configural processing has also been observed for the processing of temporal changes in dynamic emotional expressions (Ambadar, Schooler, & Cohn, 2005; Bould & Morris, 2008; Tobin, Favelle, & Palermo, 2016). Thus, configural information, for example related to changes in AU positions (Martinez, 2017), is used in the processing (e.g., accurate categorization) of dynamic facial emotion expressions.

Just as configural processing may ease the ability to detect deviations and lead to uncanniness, so may configural deviations in dynamic facial emotion expressions be perceived as uncanny. This may be the case even when the static structural configuration of the face remains unaltered, and deviations occur only on a dynamic dimension: Specifically, manipulating the sequence or synchrony of face AU movements. Asynchronous facial movement, understood here as a deviation from the typical sequence of face AU movements for a specific expression recruits different neural activity analogous to research on static face configuration processing (Skiba & Vuilleumier, 2020), and inverting a face decreases the ability to detect subtle asynchronies in dynamic expressions (Johnston, Brown, & Elson, 2021), indicating that configural processing eases the ability to detect asynchronies. Asynchronous face AU motion may thus be considered as a deviation from the typical configural pattern of a dynamic facial expression, and according to the *deviation from specialization* hypothesis, should be a source of uncanniness.

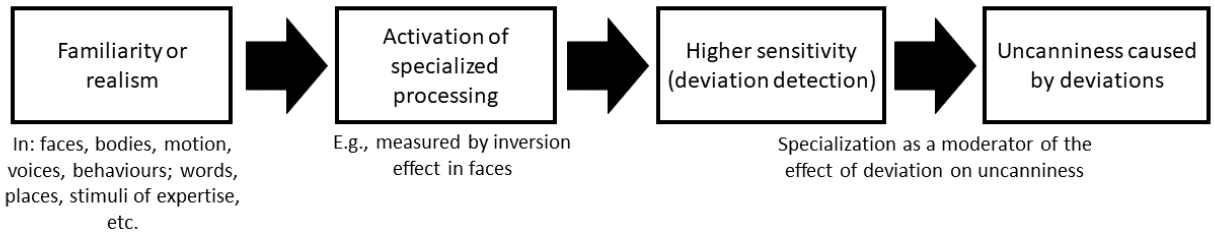
A refined theory of the uncanny

In summary, deviation from predicted pattern elicits aversion due to processes like processing disfluency or expectation violation. Familiarity or specialization with a category (e.g., faces, bodies, voices, written text, places, emotion expressions and motion) increases the number of dimensions on which a stimulus may deviate, making errors more apparent and increasing the aversion relative to a typical stimulus. As social stimuli or stimuli related to human appearance have a relatively high degree of specialization, errors created by designing an

artificial humanoid entity become more apparent, especially if the entity elicits specialized processing due to its high level of realism. The theory can be summarized as in *Figure 1.3*:

Figure 1.3

Flowchart of the refined theory of the uncanny showing its proposed underlying cognitive mechanisms.



This refined theory would extend the uncanny valley model by adding the following predictions:

1. Increasing familiarity or realism of a base stimulus increases the sensitivity to deviations (e.g., the ability to differentiate stimuli based on slight changes) and the uncanniness caused by deviations.
2. Decreasing specialized processing (e.g., inversion in faces) decreases the sensitivity to deviations and the uncanniness caused by deviations.
3. A moderated linear function of uncanniness, deviation, and specialization (e.g., inversion effect in faces) describes the data better than a polynomial "uncanny valley" function, despite a polynomial function being significant.
4. Increasing specialized processing (e.g., expertise training) increases the uncanniness caused by deviations.
5. The effects of deviation on uncanniness and the moderating effect of specialized processing are not exclusive to facial structure and can be generalized onto other

human stimulus categories (e.g., facial expressions, voices, bodies, biological motion) and inanimate categories that show evidence for specialized processing (e.g., written text, physical places).

6. Aversion to deviancy in simple patterns (deviancy aversion) predicts the uncanniness caused by deviations in faces.
7. If processing disfluency causes the aversion, then an increased neural response related to the processing of the stimulus should predict uncanniness of deviating stimuli. If expectation violation causes the aversion, then an increased neural response related to prediction errors (N400 components) should predict uncanniness of deviating stimuli.

The goal of this thesis is to critically investigate the predictions above in a series of experiments. This way, central predictions of the described refined theory of the uncanny are going to be tested (*Figure 2*). In addition, several other predominant theories and predictions of the uncanny valley will be tested, among them category-based explanations, dehumanization, disease avoidance, misattribution, mortality salience, and novelty avoidance.

Scope of the work

In sum, it is mainly investigated whether markers of specialization increase the effect of deviation on uncanniness across stimulus categories. Although a mathematical model has been proposed in this Chapter, the exact mathematical predictions (e.g., higher beta values for deviation effects on uncanniness for more specialized categories) are not investigated. Instead, the model's conceptual predictions are investigated. In addition, conceptual questions regarding the exact variables (e.g., uncanniness) or the nature of the "uncanny feeling" are not investigated as they lie beyond the scope of this work. Finally, the present research is mainly conducted with (young adult) populations from Western or Japanese

countries, and cross-cultural or cross-age effects beyond these demographic ranges are not within the scope of the work.

Research plan

The dissertation's core work is divided into ten research chapters, with each chapter containing one research manuscript that, by the time of this dissertation's submission, is either published or under review. A summary overview is provided here.

Definitions of common terms across experiments

The term *deviation* is used to describe a statistically unusual appearance of an object relative to the typical appearance within its category. When artificially edited, deviations are created by gradually changing the position of the relative features. The term deviation here will usually refer to structural deformations of a configural nature. The role of configural processing is here predominantly investigated as a form of specialized processing, and measured via the *inversion effect*: this term will refer to improved recognition abilities for upright compared to inverted stimuli. *Uncanniness sensitivity* is used to describe an increase of uncanniness compared to the (non-deviating) baseline stimulus given the same deviation manipulation. A higher uncanniness sensitivity thus refers to higher uncanniness ratings relative to the baseline given the same level of stimulus distortion (e.g., upright faces would have a higher uncanniness sensitivity than inverted faces if the difference in uncanniness between distorted and non-distorted stimuli is higher for upright compared to inverted faces). Meanwhile, the term *deviation sensitivity* (relevant only in Chapter 2) refers to the ability to detect distortions given the same level of stimulus manipulation. Finally, the term *uncanniness inversion effect* is used to describe a decrease in uncanniness following stimulus inversion, and is used as an indicator of the role of configural processing in uncanniness evaluations of a stimulus.

Expertise increases deviation sensitivity and uncanniness sensitivity

Chapters 2 to 4 investigate the role of specialized processing as a correlate to the sensitivity to structural deviations in faces (Chapters 2 and 3), and as a causal factor of uncanniness sensitivity in objects (Chapter 4).

Chapter 2 focuses on the roles of subjective familiarity with faces and configural processing on both detection sensitivity and uncanniness sensitivity: It is investigated whether an uncanniness inversion effect occurs, and whether it is mediated by a higher deviation sensitivity for upright compared to inverted faces. In addition, it is investigated whether uncanniness sensitivity is higher for familiar faces, again mediated by deviation sensitivity. In Chapter 3, it is investigated whether the inversion effect (higher recognition performance for upright compared to inverted faces) differs among different face categories classified based on their realism level, and whether this inversion effect predicts uncanniness sensitivity throughout the categories. In Chapter 4, the causal role of specialization is investigated through an expertise training paradigm: It is tested whether an expertise manipulation through training increases the uncanniness of deviating exemplars in an otherwise novel category.

In summary, Chapters 2 to 4 will provide the foundation of the refined theory (*Figure 2*): it is investigated whether specialized processing sensitizes the uncanniness of deviating stimuli, and whether this effect is mediated by a higher level of sensitivity to deviations.

Deviation sensitivity in inanimate categories

Chapters 5 and 6 focus on the application of the refined theory in inanimate categories: written text (Chapter 5) and physical places and architecture (Chapter 6). Uncanniness in inanimate stimulus categories will support stimulus-independent (e.g., cognitive) theories on the uncanny valley over those specific to a human likeness dimension (e.g., dehumanization, mortality salience, threat to human identity), or general animacy (disease avoidance).

Chapter 5 investigates the role of language familiarity (which is associated with specialized, holistic processing of words) on the uncanniness sensitivity for distorted written text, and the role of processing disfluency theories in explaining the uncanniness of orthographic distortions and semantic ambiguity. Chapter 6 investigates uncanniness of structural distortions in physical places and architectural structures, using 1) naturally occurring places that have been colloquially described as uncanny (*liminal spaces*), and 2) manipulated distortions in physical places. Together, Chapters 5 and 6 provide empirical support for the refined theory by applying it to a wider range of stimulus categories, and also extending the uncanny valley effect to previously ignored stimuli (e.g., physical places).

Critical investigation of theories

Chapters 7 to 9 provide critical investigations on various theories of the uncanny valley. *Critical investigation* here refers to research designs created with the purpose of finding falsifying results for predictions of uncanniness according to different theories. As falsification is a fundamental principle of the scientific process by weeding out insufficient explanations, and given the wide range of proposed theories on the uncanny valley, a falsification-focused approach can improve uncanny valley research by eliminating (or at least forcing to improve) theories incapable of explaining contradicting data.

Chapter 7 replicates an uncanny valley in voice stimuli while testing a variety of theories: Other than the refined theory proposed here, categorization-based theories, animacy and mind attribution, and disease avoidance theories are investigated. Specifically, it is tested whether artificial and organic deviations in voices elicit uncanniness, and whether this is mediated by categorical ambiguity. Furthermore, it is investigated whether uncanny voice categories also elicit contradicting attributions of mind and animacy.

Chapter 8 tests the refined theory by investigating the uncanniness inversion effect in ecologically valid stimuli related to the uncanny valley: video clips of androids or CG characters deemed uncanny in previous research. Furthermore, an emotional priming paradigm and a lexical decision task (LDT) are applied to test whether an uncanny prime elicits changes in LDT reaction times analogous to disgust primes and fear primes in the processing of disease- or death-related words, as would be expected by disease avoidance and mortality salience theories.

Chapter 9 presents a neurophysiological investigation of two theories on the uncanny valley: expectation violation and processing (dis-)fluency. As the results from the previous chapters can be explained by both theories, a critical investigation of both theories is lacking. Previous research can establish different event-related potential (ERP) components for each theory: While location-general N400 components have been associated with prediction errors (Urgen et al., 2018), increased amplitudes for location-specific components, such as P100 and N170 for faces, can be associated with increased processing need potentially caused by disfluency (Olivares et al., 2015). Chapter 9 investigates these two competing theories through theory-specific neurophysiological markers correlating with presentation of uncanny stimuli.

Chapter 10 investigates individual differences in various traits as predictors of uncanniness across previously tested uncanny deviating stimuli of different object categories (androids, clowns, bodies, faces, places, voices, written text). Measured trait constructs are deviancy aversion, disgust sensitivity, neuroticism (anxiety facet), and individual need for structure. In addition, as previous research suggested a link between the uncanny valley and coulrophobia (Tyson et al., 2023), it is investigated whether clowns fall into an uncanny valley and whether trait coulrophobia is associated with the uncanny valley.

In sum, Chapter 7 to 10 extend uncanny valley research by critically investigating various theories on the uncanny valley, including the refined theory proposed in this dissertation.

Application in a humanlike android

In Chapter 11, the refined theory is applied for the investigation on the uncanny valley using facial emotion expressions of a realistic android Nikola (Sato et al., 2022). Chapter 11 not only extends the refined theory to include deviations in dynamic expressions, but is also used to investigate the uncanniness inversion effect in a realistic humanlike android. Thus, Chapter 11 serves as a final verification in the refined theory using an ecologically valid stimulus.

The final Chapter 12 summarizes and discusses the research and its implications and presents future directions.

Chapter 2: Familiarity, orientation, and realism increase face uncanniness by sensitizing to facial distortions

Methods, experiments and large portions of the introduction and discussion in this chapter have been published in the *Journal of Vision* (Diel & Lewis, 2022a).

Potential links between the UV effect and face-related processing have been suggested, like configural processing or perceptual narrowing (Almaraz, 2017; Diel & MacDorman, 2021; Kätsyri, 2018; MacDorman & Chattopadhyay, 2017). The link between configural processing and the UV was investigated here. Specifically, it was investigated whether correlates of specialization in faces (familiarity, upright orientation, realism) increase the detection to subtle changes in a face, which in turn increase negative evaluation.

Uncanny valley and face processing

Humans are specialized for natural human faces (Kanwisher, 2000). Specialized processing may sensitize the detection and devaluation of subtle distortions, leading to a UV effect for realistic stimuli. While individuals show a remarkable ability to recognize faces, this ability is reduced for virtual faces, indicating that face expertise does not transfer to computer-generated faces (Crookes, et al., 2015). Kätsyri (2018) had participants learn and later recognize and rate a set of real and virtual faces and found a higher false alarm rate for recognizing virtual compared to real faces, again indicating difficulties in differentiating virtual faces when compared to real ones. Furthermore, inversion increased the eeriness of both virtual and real faces and more so for real ones, which Kätsyri (2018) argued to be evidence against the role of configural processing on the uncanniness of faces. However, previous research has also shown that inversion reduces the variation of aesthetic judgments of faces (Bäuml, 1994; Leder et al., 2017; Santos & Young, 2008). Thus, configural information may instead be used to accurately assess facial aesthetics, for example, subtle configural deviations may appear less attractive or more eerie. Configural processing would then increase the range of aesthetic ratings across different face configurations due to a higher sensitivity to configural variation, including the difference between real and virtual faces. Although inversion itself may increase face eeriness in general because upside-down faces are more atypical than upright faces, inversion would then also decrease the effect of distortions on the variance of aesthetic ratings due to the decrease of perceived configural variance, and reduce the eeriness difference between real and virtual faces. Thus, it is possible that Kätysri's (2018) observation that the eeriness difference between real and virtual faces decreased when inverted may have resulted from the decreased ability to detect configural information.

Kätsyri (2018) did not manipulate the degree of face distortion, whereas the presumed moderating effect of inversion on the uncanniness of face distortions should be especially salient with a wider range of face distortions and especially for highly distorted faces. Specifically, inversion should lessen the increase of uncanniness across incremental facial configural distortions. In other terms, inversion should attenuate the effect of configural deviations on uncanniness by decreasing perceptual sensitivity to these deviations.

A higher level of face realism enhances sensitivity of the uncanniness of facial distortions (MacDorman et al., 2009; Mäkäräinen et al., 2014). Matsuda, Okamoto, Ida, Okanoya, and Myowa-Yamakoshi (2012) have furthermore suggested that a high degree of perceptual expertise for a face would also increase the sensitivity to deviations and fine-detail errors within the face. More generally, increased expertise or familiarity would translate into higher distortion sensitivity, and thus a stronger UV effect for humanlike compared to non-humanlike categories (e.g., distorted human compared to animal faces). Similarly, if perceptual familiarity drives the ability to detect subtle deviations, a higher distortion sensitivity and UV effect would be expected for familiar compared to novel faces. This proposal, summarized as *deviation from familiarity hypothesis*, has not yet been investigated in previous research.

Experiment 1

Research questions and hypotheses

Experiment 1 aimed to investigate the effect of face familiarity and inversion when interacting with the level of facial distortion, on two variables: (1) uncanniness ratings of faces, and (2) the ability to detect changes in facial distortion (*distortion sensitivity*). Previous researchers proposed that a high level of perceptual expertise leads to perceptions of uncanniness caused by improved detection of subtle configural distortions (e.g., Matsuda et al., 2012). Thus, uncanniness ratings should increase with increasing facial distortion

(distortion main effect). This effect should be stronger for familiar (compared to novel; distortion-familiarity interaction) and upright (compared to inverted; distortion-orientation interaction) faces given the higher specialization with both familiar and upright faces.

Second, the ability to detect changes between two variants of a same face (e.g., a normal face and a slightly distorted version) should increase with a higher level of distortion difference between the faces (distortion main effect). This distortion difference level should interact with both familiarity (higher distortion sensitivity for familiar compared to novel faces) and orientation (higher distortion sensitivity for upright compared to inverted faces) if familiarity enhances the ability to detect distortions.

Finally, if uncanniness is caused by the ability to detect distortions, distortion sensitivity, here, operationalized as the degree of distortion necessary to accurately differentiate between distorted versions of the same face, should predict the sensitivity of uncanniness across different face conditions. Thus, the hypotheses are the following:

- Both face familiarity and face orientation interact with face distortion on the effect of uncanniness: familiar and upright faces show a stronger increase for uncanniness ratings with increasing the distortion levels compared to unfamiliar and inverted faces.
- Both face familiarity and face orientation interact with face distortion on the effect of distortion sensitivity: familiar and upright faces have a higher distortion sensitivity than unfamiliar and inverted faces.
- Distortion sensitivity predicts the effects of familiarity, orientation, and face distortion on uncanniness ratings.

Rating scales are the preferred method of measuring uncanniness in UV research, as they allow measuring a differentiated subjective experience (Diel et al., 2022; Ho & MacDorman,

2017). For stimulus ratings, some of the most used and most effective ratings scales in uncanny valley research were used according to a meta-analysis (Diel et al., 2022): *creepy*, *eerie*, *repulsive*, and *strange*. Items were combined into an uncanniness index. In addition, human likeness was measured with a single scale. To measure distortion sensitivity, a two-back delayed face matching to sample task was used, a setup used in previous face differentiation studies (e.g., Rhodes, Hayward, & Winkler, 2006).

Methods

***Participants*.** Sixty-six participants took part in the experiment. Thirty-three British participants were recruited via the Cardiff University School of Psychology's Experimental Management System (EMS; M_{age} = 19.15, SD_{age} = 1.56), and 33 German participants were recruited via Prolific (M_{age} = 24.73, SD_{age} = 3.52). Participants either received course credits or a small monetary reward for participation.

Stimuli. In a preliminary study, images of 28 individuals were collected depicting frontal faces of 14 famous British and 14 famous German persons. All face stimuli were cropped to equal size, coloured, and only showed the head, ears, neck, and parts of the hair. Facial expressions were either neutral or, if no neutral expression of the individual was obtainable, happy. Twenty British and 20 German participants were asked to rate whether they recognized each face and, if so, to state either the name of the person or the context the in which the person appears. The number of correct recognitions were counted for British and German participants. The five British and five German faces that were recognized most often by participants from the same country while recognized least often by participants from the other country were selected as stimuli for the main experiment. The famous British faces and the number of times recognized by the British and German participants were Philipp Schofield (33 British and 2 German), Holly Willboughly (33 British and 1 German), Anthony McPartlin (31 British and 1 German), Rylan Clark-Neal (31 British and 0 German), and Gary

Lineker (21 British and 0 German). Famous German faces and the number of times recognized were Dieter Bohlen (0 British and 33 German), Thomas Gottschalk (0 British and 33 German), Stefan Raab (0 British and 32 German), Günther Jauch (0 British and 28 German), and Otto Waalkes (0 British and 26 German).

Photographs of those 10 selected famous persons were used as test stimuli. Each face was distorted in standardized steps by incrementally increasing the distance between the eyes while lowering the mouth. For each distortion level, interocular distance was increased by laterally displacing eyes so that the medial border of each eye's iris is placed between its original position and the position of the eye's pupil of the previous distortion level. The mouth was moved toward the chin to position the upper vermilion border between its original position and the oral fissure of the previous distortion level. Each face had five variations of incrementally increasing distortions, including the original face. Here, the term *face identity* is used to refer to an identity depicted by the face regardless of the face's distortion level, and the term *base face* to refer to the original, unedited face. Finally, all face variants were inverted on the horizontal axis to create two orientation conditions (upright and inverted). Figure 2.1 shows the distortion variations one example face. Face stimuli were edited using the Photoshop CS6 software.

Figure 2.1

An illustration of the five examples of face stimulus distortion levels. Note. Faces were also presented inverted. The face depicted was not used in the experiment. The face was artificially created by the StyleGAN generative network (Karras, Laine, Aittala, Hellsten, Lehrinen, & Aila, 2020).

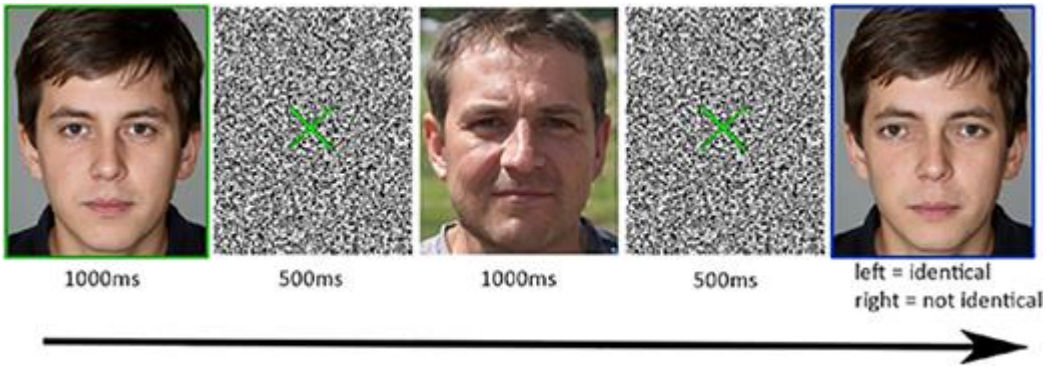
Face rating task. The first task consisted of rating each of the 100 faces (2 orientation × 5 distortion levels, for 10 face identities) on five scales: *eerie* (*unheimlich*), *creepy* (*gruselig*), *strange* (*merkwürdig*), *repulsive* (*abstoßend*), und *humanlike* (*menschenähnlich*). Each scale ranged from 0 (not at all) to 100 (fully) and were presented in the language preferred by the participant (English or German). Faces were presented randomly, and, for each face, scales were presented sequentially, simultaneously with the face. Participants had unlimited time to view the face and select a response.

Delayed face matching to sample task. In the second task, a cue face (surrounded by a green square) was presented followed by grey noise with a green fixation cross, a distractor (masking) face, again noise/cross, and a match face (surrounded by a blue square). Cue and match faces were always variations of the same face identity, presented in the same orientation, and were of the same or different distortion levels. Participants had unlimited time to view the match face and to decide whether the match face exactly matched the cue face. Participants had to press the left arrow key to indicate that the faces were identical, and the right arrow key to indicate that they were not identical. Masking stimuli were faces of other famous British or German persons that were not used as test stimuli in this experiment. A single trial is depicted in *Figure 2.2*.

Figure 2.2

A trial in the delayed face matching to sample task. This is a mismatched trial as cue (green surround) and target (blue surround) faces are not identical. Note. The example faces were

not used in the actual experiment. The faces were artificially created by the StyleGAN generative network (Karras et al., 2020).



All distortion levels of a face were matched with one another, combining into 25 cue-match face pairs per face identity. Given 2 × 10 different base faces (orientation × famous person), the task consisted of a total of 500 trials where each face pair was shown once while each face appeared five times. Faces were identical 20% of the time. A break was offered every 50 trials.

Procedure. The study was conducted online. After receiving the link to the study, participants consented to the experiment and filled a short demographic questionnaire and a questionnaire on whether participants could recognize and identify each of the 10 famous persons. The response was used to control familiarity in the experiments. Participants then completed the face rating task first and the delayed face matching to sample task second. Because the exposure to each face was higher in the matching compared to the rating task, the rating task was conducted first to reduce the effect of familiarization on the experiments. After the study, participants received a debriefing.

Statistical analysis. For the first hypothesis, a 2 × 2 × 5 (orientation × familiarity × distortion level) analysis was conducted for uncanniness ratings, with orientation, familiarity, and distortion level as fixed effects and face identities and participants as random effects. For the second hypothesis, a 2 × 2 × 5 (orientation × familiarity × distortion difference level) analysis

was conducted for "identical" response rate, with orientation, familiarity, and distortion difference level as fixed effects and face identities as random effects. For the third hypothesis, "identical" response rates were added as a fixed-effect predictor for the model used for hypothesis 1. Data cleaning was conducted by removing all interquartile range outliers for each distortion condition (distortion levels 0 to 4). Data preparation, data cleaning, and statistical analyses were conducted in R software. Linear mixed models[1] were used for hypotheses 1 to 3 because they allow to deal with both fixed effects and random effects (McLean, Sanders, & Stroup, 1991), which are expected in the present study given the within-subject and within-face design. Linear mixed models are more appropriate than standard ANOVA here because of the need to control for the effect of face identity. This type of analysis produces the large degrees of freedom that can be observed below (see also Kuznetsova, Brockhoff, & Christensen, 2017; Luke, 2017). The R software packages *lme4* (for linear mixed models, using the function *lmer()*) and *lmerTest* (for complete depiction of the results), and *robustlmm* were used (Bates, Mächler, Bolker, & Walker, 2015).

Ethics statement and data availability. The study was approved by the Cardiff University School of Psychology Research Ethics Committee in in November 2020 (reference number: EC.20.10.13.6081GR). The data and the R code for the analysis are available at: https://osf.io/7prax.

[1] Linear mixed models allow to statistically control for random effects caused by groupings of the data. In this case, initial differences in the aesthetic ratings between the faces' identities may distort the data if face identity is not controlled for. Linear mixed models are used throughout the dissertation.

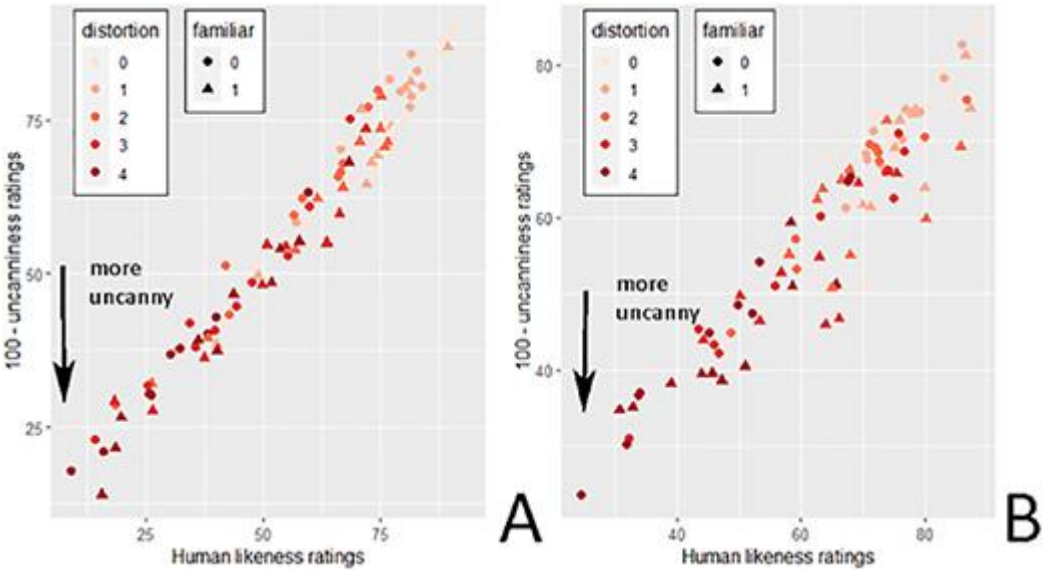
Results

Scale evaluation. The scales eerie, creepy, strange, and repulsive were combined into a single uncanniness index by calculating the mean values across the four scales after correcting scale inversions. The index' Cronbach's alpha was $\alpha = 0.94$, indicating strong reliability.

Uncanniness and human likeness. Uncanniness ratings were plotted as a function of human likeness. A linear mixed model could explain the distribution ($t(502) = -81.22$, $p < 0.001$), whereas a quadratic model could not ($t(5108) = -3.021$, $p = 0.239$). Thus, the relationship between uncanniness and human likeness is best explained by a linear function. *Figure 2.3* shows a scatterplot with each point depicting a trial, for both upright and inverted faces.

Figure 2.3

*Uncanniness ratings as a function of human likeness ratings for (**A**) upright and (**B**) inverted faces, across distortions levels (0 = base face) and face familiarity. The "100- uncanniness ratings" represent the y-axis of Mori's (2012) original uncanny valley curve, with lower values depicting higher uncanniness ratings.*

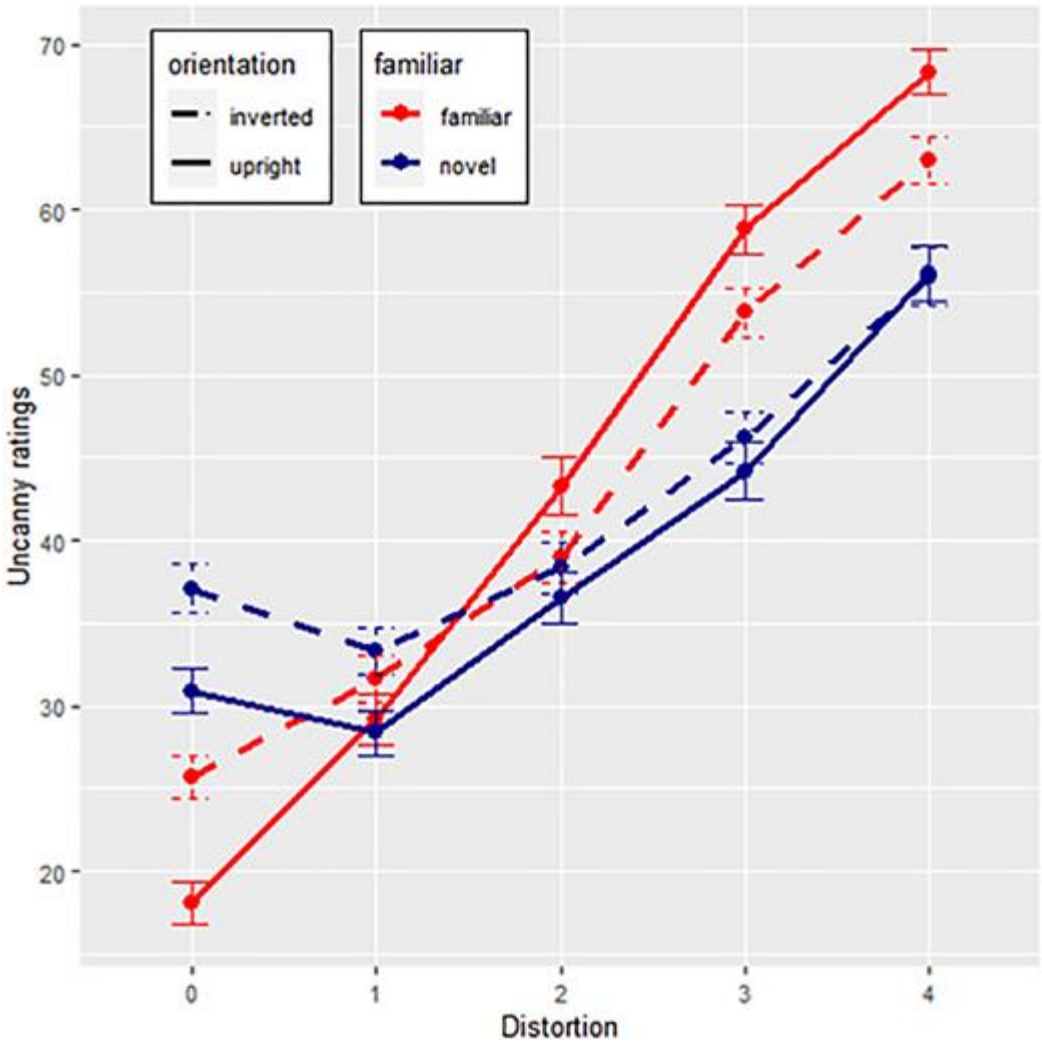


Post hoc linear mixed model analyses found that human likeness ratings decreased with increasing distortion levels ($t(5037) = -29.551$, $p < 0.001$) and that novel faces were more humanlike than familiar faces ($t(5055) = 3.24$, $p = 0.001$). However, no main effect of orientation was observed ($t(18) = 0.871$, $p = 0.395$). Distortion interacted with both familiarity ($t(5037) = 5.678$, $p < 0.001$) and orientation ($t(5037) = 10.194$, $p < 0.001$).

Prediction of uncanniness. Orientation, familiarity, and distortion were used as fixed effects to predict uncanniness, and face identity and participants as random effects. As the assumption of homoscedasticity was not met, a robust estimation of the linear mixed model was calculated. Distortion significantly predicted uncanniness ($t(6308) = 32.483$, $p < 0.001$), but neither familiarity ($t(6317) = 0.257$, $p = 0.798$) nor orientation ($t(19) = -1.073$, $p = 0.297$). Interaction effects between distortion and familiarity ($t(6308) = -6.204$, $p < 0.001$), distortion and orientation ($t(6308) = -11.573$, $p < 0.001$), and familiarity and orientation ($t(6321) = 2.644$, $p = 0.008$) were found, as well as an interaction with all factors combined ($t(6308) = 2.588$, $p = 0.010$). The model's regression coefficient was $R^2_{corr} = 0.458$. Data are summarized in *Figure 2.4*.

Figure 2.4

Uncanny ratings across face distortion levels (0 = original face, 4 = most distorted face). Red and blue lines depict ratings for familiar and unfamiliar faces, whereas slashed and full lines depict response rates for inverted or upright faces. Error bars show +/−1 standard errors based on within-subject variability.



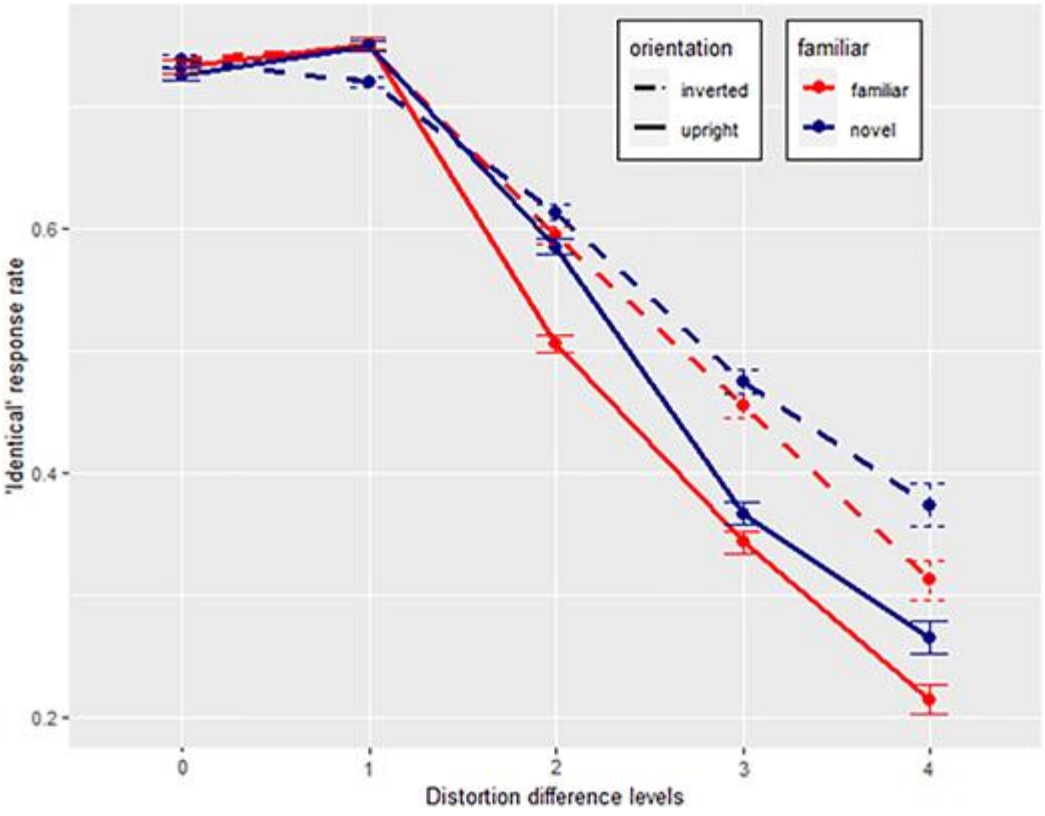
Post hoc Tukey tests with Bonferroni corrections were performed to test differences between face condition groups (familiar upright versus novel upright, familiar upright versus familiar inverted, novel upright versus novel inverted, and familiar inverted versus novel inverted) for each distortion level. At distortion level 0, novel upright faces were more uncanny than familiar upright faces ($t(65) = 4.657$, $p_{adj} < 0.001$), familiar inverted faces were more uncanny than familiar upright faces ($t(65) = 6.324$, $p_{adj} < 0.001$), and novel inverted faces more uncanny than familiar inverted faces ($t(65) = 2.748$, $p_{adj} = 0.031$) and novel upright faces ($t(65) = 5.103$, $p_{adj} = < 0.001$). Thus, both novelty and inverted orientation increased uncanniness of base faces. At distortion level 1, no differences between condition groups were significant. At distortion level 2, all differences were nonsignificant except for familiar

inverted faces, which were less uncanny than familiar upright faces ($t(65) = -4.482$), $p_{adj} = < 0.001$). Thus, at distortion level 2, upright orientation increased uncanniness ratings for familiar faces. Familiar inverted faces remain less uncanny than familiar upright faces at distortion level 3 ($t(65) = -8.47$, $p_{adj} < 0.001$), and novel inverted faces become less uncanny than familiar inverted faces ($t(65) = -4.331$, $p_{adj} < 0.001$). Thus, at this stage, inversion generally reduces the uncanniness of distorted faces. Finally, at distortion level 4, familiar inverted faces again remain less uncanny than familiar upright ($t(65) = -8.072$, $p_{adj} < 0.001$), and novel inverted faces less uncanny than familiar inverted faces ($t(65) = -4.727$, $p_{adj} < 0.001$). In addition, novel upright faces are less uncanny than normal familiar faces ($t(65) = -2.963$, $p_{adj} = 0.023$), suggesting that both upright orientation and familiarity increase the uncanniness of distorted faces. These results show that uncanniness increases the strongest across distortion levels when faces are upright (versus inverted) and familiar (versus novel). Thus, hypothesis 1 was supported.

Face matching task. All participants with an "identical" response rate of equal or less than 25 between distortion difference levels 0 and 4 were excluded, as no difference in response behavior between the end point distortion levels indicates that participants answered at random. A total of 16 data sets were excluded from the response rate analysis, leaving $n = 50$ participants (21 British and 29 German). Data are summarized in *Figure 2.5*.

Figure 2.5

Identical response rates across face distortion difference levels (0 = cue and match face were identical, 4 = cue and match face were 4 distortion levels apart) Red and blue lines depict familiar and unfamiliar faces, whereas dashed and full lines depict response rates for inverted and upright faces. Error bars show +/−1 standard errors based on within-subject variability. Note. Distribution bars represent standard deviations.



Orientation, familiarity, and distortion difference level were included as fixed effects to predict identical response rates, and face identity, and participant as random effects. The assumption of homoscedasticity was not met, hence, a robust estimation of the linear mixed model was performed. Distortion difference levels ($t(24900) = -65.097$, $p < 0.001$), familiarity ($t(24910) = 10.996$, $p < 0.001$), and orientation ($t(24370) = 16.853$, $p < 0.001$) all significantly predicted identical response rates, just as interactions between distortion and familiarity ($t(24900) = 5.419$, $p < 0.001$), distortion and orientation ($t(24900) = 10.707$, $p < 0.001$), and familiarity and orientation ($t(24910) = -4.989$, $p < 0.001$). The model's regression coefficient is $R^2_{corr} = 0.449$.

Post hoc Tukey tests were conducted to test differences between condition groups (familiar upright versus familiar inverted, familiar upright versus novel upright, novel upright versus novel inverted, and familiar inverted versus novel inverted) across distortion difference

levels. At distortion difference levels 0 and 1, no tested differences were significant. At distortion difference level 2, only familiar inverted faces had a higher identical response rate than familiar upright faces ($t(41) = 3.559$, $p_{adj} = 0.007$). Thus, at distortion difference level 2, familiar faces were easier to discriminate when they were upright compared to inverted. At distortion difference level 3, familiar inverted faces remained more difficult to differentiate than familiar upright faces ($t(65) = 3.618$, $p_{adj} = 0.006$), in addition to novel inverted faces having a higher identical response rate than novel upright faces ($t(65) = 3.441$, $p_{adj} = 0.01$). Thus, inversion decreased the general ability to differentiate between faces at this distortion difference level. Finally, at distortion difference level 4, familiar inverted faces had still a higher identical response rate than familiar upright faces ($t(65) = 3.39$, $p_{adj} = 0.006$) and novel inverted faces higher than novel upright faces ($t(65) = 3.441$, $p_{adj} = 0.01$). In addition, novel inverted faces had a higher identical response rate than familiar inverted faces ($t(65) = 2.206$, $p_{adj} = 0.016$), but the difference between familiar and novel upright faces remained nonsignificant. Identical response rate decreased stronger for upright faces across distortion levels than for inverted faces, especially when faces were familiar. Thus, hypothesis 2 is supported.

Distortion sensitivity as a predictor of uncanniness. According to the third hypothesis, the ability to detect distortion differences of the same face can explain the effects of familiarity, orientation, and distortion on uncanniness. Thus, the rate of identical responses (study 2) was added to the prediction model of uncanniness (study 1). Because the variables from the two studies were coded differently (uncanniness ratings are linked to individual stimuli, whereas response rates are linked to pairs of two stimuli), only study 2 trials with a base face as either cue or target face were included, and response rates were linked to the uncanniness ratings of the face paired with the base face (or of the base face if cue and target were identical).

A linear mixed model was calculated either with identical response rate, or familiarity, orientation, and distortion as fixed effects and face identity and participants as random effects. Because the assumption of homoscedasticity was not met, robust estimations were calculated. Significant main effects for all predictors (for familiarity $t(9388) = 6.684$, $p < 0.001$; for orientation $t(9120) = -11.077$, $p < 0.001$; for distortion $t(9386) = 34.314$, $p < 0.001$; and for response rate $t(9340) = 2.232$, $p = 0.026$) were found. Furthermore (and in correspondence to the previous regression analyses), the interactions between familiarity and orientation ($t(9385) = 7.029$, $p < 0.001$), distortion and familiarity ($t(9380) = -4.819$, $p < 0.001$), distortion and orientation ($t(9281) = -11.051$), $p < 0.001$), and distortion, familiarity, and orientation combined were significant ($t(9378) = 3.702$, $p < 0.001$). The interactions remain significant when adding the identical response rate as a predictor (for familiarity, orientation, and response rate $t(9382) = 2.188$, $p = 0.029$; for distortion, familiarity, and response rate $t(9381) = -5.736$, $p < 0.001$; for distortion, orientation, and response rate $t(9382) = -6.900$, $p < 0.001$; and for all predictors combined $t(9379) = 3.348$, $p < 0.001$). The model's regression coefficient is $R^2_{corr} = 0.511$.

A model with response rate alone could predict uncanniness ratings ($t(9431) = -38.37$, $p < 0.001$, $R^2 = 0.371$). The three factors of orientation, familiarity, and distortion could predict the response rate, with an R^2 of 0.451. Thus, hypothesis 3 was supported.

Discussion

Human likeness ratings. The results show a linear relationship between human likeness and uncanniness. As realistic faces and their distortions were used in this study and no less humanlike stimuli, the results are not surprising: the stimulus range and data likely reflect the rightmost part of the valley or the range from the low point of the valley to full human likeness. Post hoc analyses found interaction effects between the distortion level and the face orientation and familiarity. Specifically, human likeness decreased stronger with increasing

distortion levels when faces were upright (versus inverted) and familiar (versus novel). These findings reflect those of uncanniness ratings: upright orientation and familiarity increase the sensitivity to human likeness perception caused by configural deviations from “normal” faces. Hence, the findings suggest that a disruption of the configural, upright face pattern also disrupts the accuracy of human likeness ratings similar to the perception of humanness in inverted faces found in previous research (Hugenberg et al., 2016).

However, an increase of uncanniness along a manipulation variable alone is not sufficient to locate a stimulus range across a “proper” UV curve because the range of human likeness to the left of the observed data is missing. Thus, additional research is needed to investigate the association between face distortion and an UV plot.

Familiarity, orientation, and uncanniness. Results show significant interactions among distortion levels, familiarity, and orientation of faces on uncanniness. Uncanniness increased across distortion levels, and this effect was reduced when faces were inverted while familiarity enhances the effect. Results thus support hypothesis 1.

Familiarity, orientation, and distortion sensitivity. In tune with hypothesis 2, familiarity and upright orientation increases the distortion sensitivity of faces. Results show significant interactions between familiarity and distortion difference and orientation and distortion difference. Specifically, both familiarity and an upright orientation increased participants’ abilities to differentiates variants of the same face.

Distortion sensitivity as a mediator for uncanniness. In accordance with previous research, stimulus categories participants are expectedly more familiar with are more sensitive to uncanniness when distorted (Chattopadhyay & MacDorman, 2016; Diel & MacDorman, 2021; MacDorman et al., 2009; Mäkäräinen et al., 2014; Matsuda et al., 2012). Perceptual experience or familiarity could affect uncanniness by increasing the viewer's ability to detect

subtle configural differences of a stimulus, thus increasing the likelihood to detect subtle deviations which are then perceived as uncanny. Although face inversion would reduce this ability because of the specialization for upright faces, familiarity would in turn enhance it. This study's results found that the response rate alone could predict uncanniness. Thus, distortion sensitivity may in fact mediate the effect of familiarity and orientation of the sensitivity to uncanniness across distortions.

Experiment 2

Although Experiment 1 found that the sensitivity to uncanniness is stronger for upright and familiar faces, the results do not allow an interpretation in the context of the UV. Whereas it is possible that the range of stimuli encompasses the rightmost part of the UV curve, this relationship has not been tested here. Furthermore, it is unclear how the degree of realism interacts with the observed effects on uncanniness sensitivity. Thus, Experiment 1 was designed to investigate whether the faces observed in Experiment 1 can be placed within a "proper" UV function, and how the level of realism interacts with familiarity and upright orientation.

Research question and hypotheses

Previous research suggests that facial distortions are more acceptable for less realistic faces (e.g., Mäkäräinen et al., 2014). This has anecdotal face value as cartoon characters are liked despite exaggerated, stylized proportions of a face or facial features which would be unacceptable for more realistic faces. One explanation is that a higher level of realism directly increases the sensitivity to deviations by decreasing the range of acceptable variation of facial structure. According to the face space framework (Valentine, 1991; Valentine, Lewis, & Hills, 2016), human faces can vary on different dimensions of facial structure. Normal variations on these dimensions which are typically observed in everyday life would create an experience-based, "acceptable" range of facial structure, whereas exaggerated

values on these face space's dimensions would lead to unusual, distorted faces places beyond this acceptable or normal range. Less realistic faces could miss important details that allow the estimation of the face's structure, which would decrease the ability to detect deviating variations and thus increase the range of acceptable face structures. Furthermore, the effect of lower realism on acceptable face variations would be more increased for inverted faces as inversion has been shown to decrease distortion sensitivity in study 1. However, face familiarity should curb the effect of low face realism on distortion, as a distorted familiar face would be judged more harshly based on its difference from the familiar norm rather than the general face norm.

Thus, the following hypotheses are proposed:

- Uncanniness of faces ranging on distortion, familiarity, orientation, and realism produce an uncanny valley-like, quadratic function when plotted against their human likeness.
- Familiarity, upright orientation, and high face realism increase the effect of distortion on uncanniness. Specifically, the increase of uncanniness across distortions is higher in more realistic faces than low realistic faces, and more so for familiar (versus novel), and upright (versus inverted) faces.

Methods

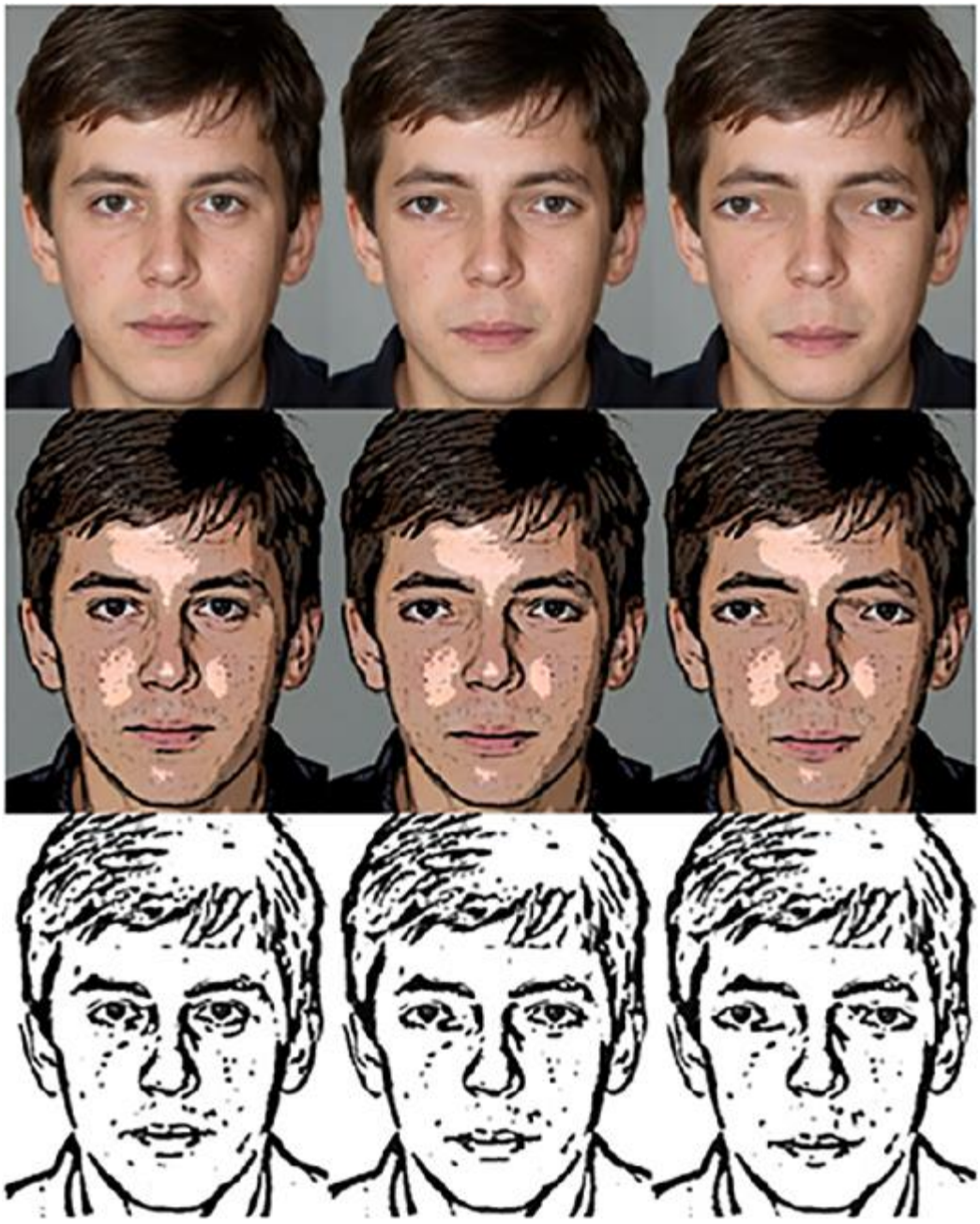
Participants. Forty-two participants have been UK participants recruited via Prolific. Participants' age was M_{age} = 24.58, SD_{age} = 4.93, and 67.5% were women.

Stimuli. Stimuli were selected and created to vary along familiarity (familiar versus upright), orientation (upright versus inverted), face distortion level (3 levels), and realism level (3 levels). First, all stimuli from study 1 with the distortion levels 0 (base face), 2, and 4 were again used in this experiment. Only three distortion levels were used to limit the total number

of stimuli. In addition, two low realism (stylized) variants of each used famous face were created using the following methods: (1) block print style, created by adding a poster edges filter to the faces in Photoshop CS6, and (2) drawing style, created by adding using the smart blur, high pass, threshold, and palette knife tools in Photoshop CS6. Image manipulation was loosely based on the method used in Mäkäräinen et al. (2014). Images consisted of five (realism) times three (distortion) times two (familiarity) times two (orientation) times five (exemplars), adding up to a total of 300 stimuli. Examples of the stylization across distortion levels are seen in Figure 2.6.

Figure 2.6

Example stimuli across distortion levels (0, 2, and 4; left to right) and realism levels (real, block print style, drawing style; and up to down). Note. Depicted example faces were not used in the actual experiment.



Because these stylized versions of famous faces can themselves be considered deviations from familiar faces, a total of 20 (2 familiarity conditions × 2 realism levels × 5 faces) faces of real cartoon characters were additionally selected and analogously distorted on three distortion levels. To control face familiarity, faces were either internationally famous or from Soviet or Russian cartoons. Furthermore, faces were either 2D- or 3D-animated to control for the level of detail. Five famous 2D animated faces were of Mickey Mouse (Disney), Homer Simpson (The Simpsons), Shaggy (Scooby Doo), Fred Flintstone (The Flintstones), and

Stewie Griffin (Family Guy). Five Soviet/Russian 2D animated faces were Uncle Fyodor (Three from Prostokvashino), Malish (Soviet animated version of Karlson from the roof), Ivan Zarevich (Ivan Zarevich and the Grey Wolf), Alyosha Popovich (Three Bogatyrs), and Jim Hawkins (Soviet animated version of Treasure Island). Five famous 3D animated faces were Super Mario (Nintendo), Elsa (Disney's Frozen), Buzz Lightyear (Disney's Toy Story), Wallace (Wallace and Grommit), and Shrek (Shrek). Soviet/Russian 3D animated faces were Masha (Masha and the Bear), Cheburashka (Cheburashka/Gena the Crocodile), the smallest gnome (samyy malenkiy gnom), Dim Dimych (Fixiki), and Boria/Valery (Fantasy Patrol). Soviet or Russian animated characters were selected because of the wide range of animated series available mostly unknown to Western audiences. All faces were either upright or inverted, creating a total of 300 faces (2 familiarity × 2 orientation × 5 realism levels × 3 distortion level × 5 faces). Selecting images of different characters or objects is one of the most common practices in uncanny valley research (see *distinct entities* in Diel et al., 2022).

Procedure. After giving informed consent, participants completed a short demographic questionnaire and followed a link to the face rating task. The face rating task was identical to the face rating task in Experiment 1.

Results

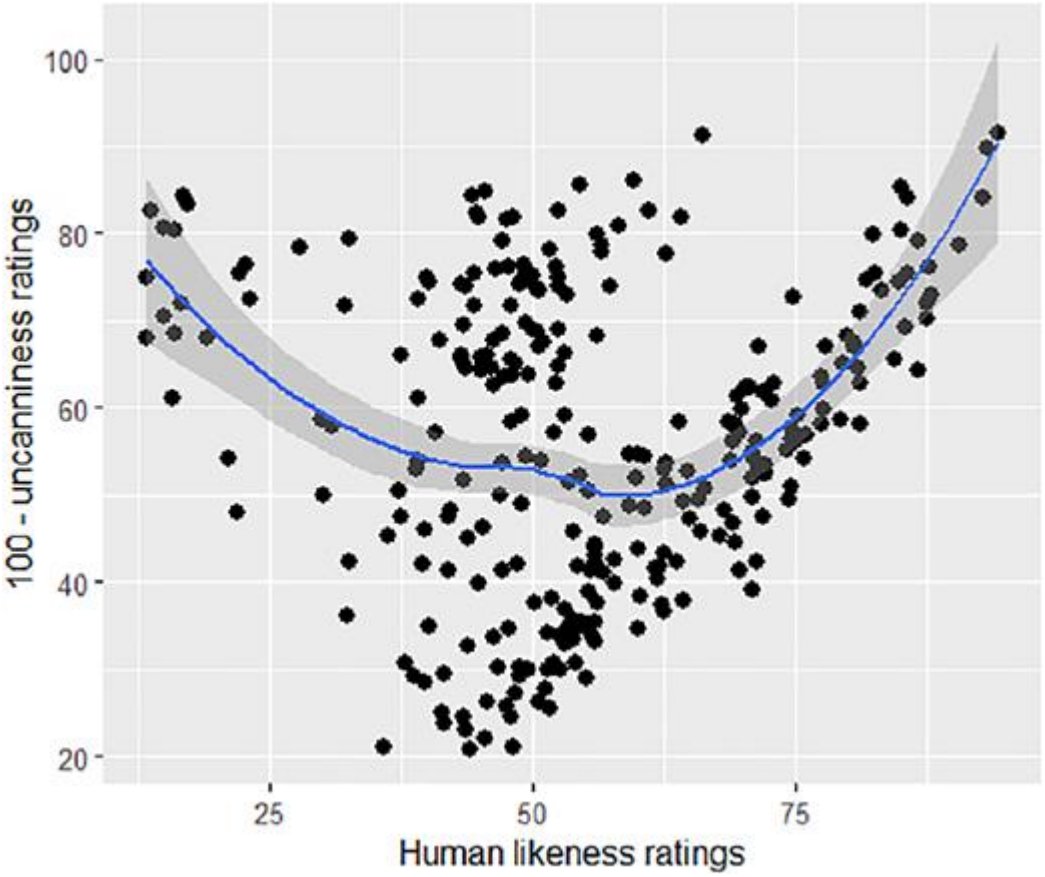
Rating scales. The scales *eerie*, *creepy*, *strange*, and *repulsive* were combined to a single *uncanniness* index. The index' Cronbach's alpha was $\alpha = 0.93$, indicating strong reliability.

Uncanny valley. To test the first hypothesis, uncanniness ratings were plotted against either linear or quadratic human likeness ratings as fixed effects in a mixed model, including base faces and participants as random effects. Both a linear function ($t(11570) = 17.45$, $p < 0.001$, $R^2_{corr} = 0.511$), and a quadratic function ($t(11560) = -30.37$, $p < 0.001$, $R^2_{corr} = 0.534$)

of human likeness were significant. The quadratic model was a better fit than the linear model ($\chi^2 = 888$, $p < 0.001$). The plot is depicted in Figure 2.7, showing an inverted U-shaped function. Thus, hypothesis 1 was supported.

Figure 2.7

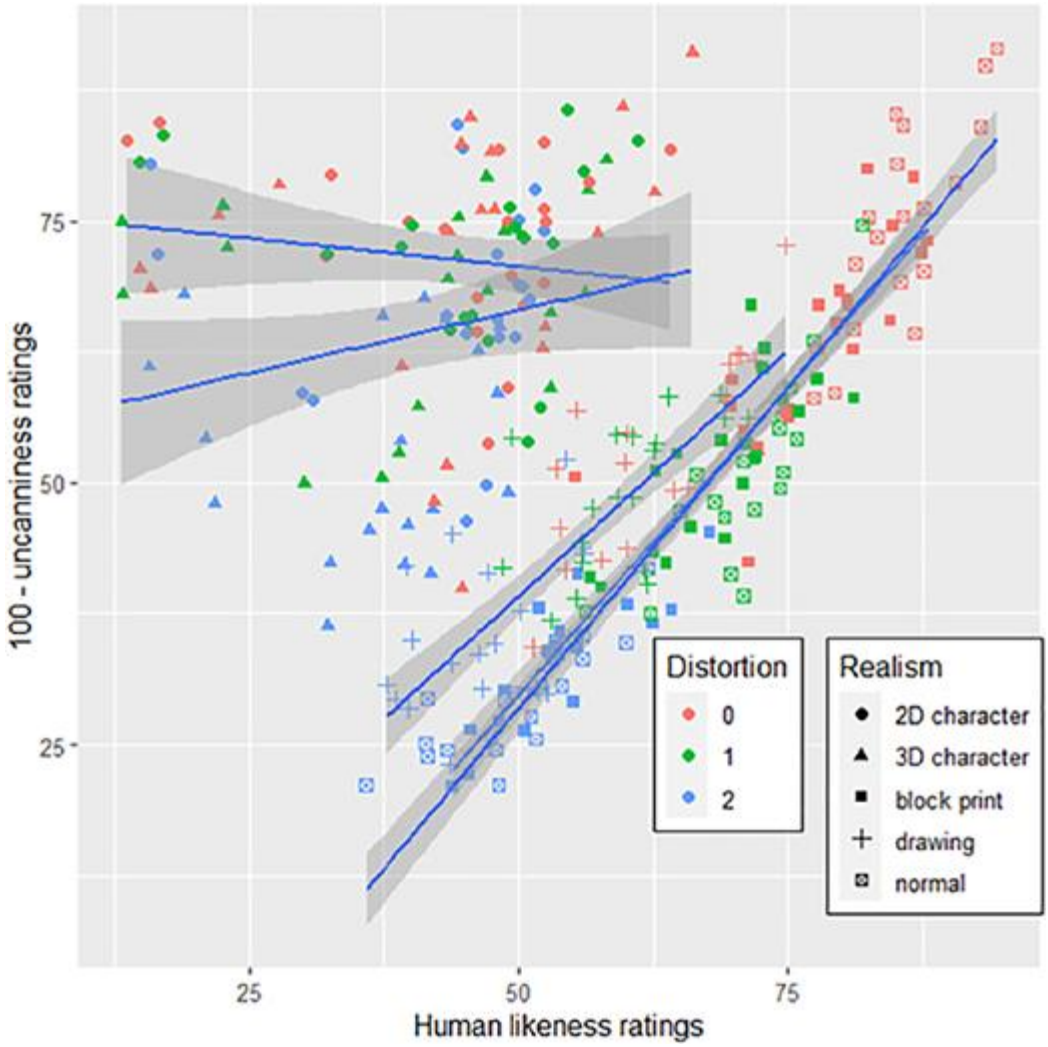
Inverted uncanniness ratings plotted against human likeness ratings. Each point corresponds to a face stimulus per condition, averaged across participants. The blue line represents the regression curve and the grey zone the confidence interval.



Furthermore, averaged fully realistic faces of all distortion levels ranged in their human likeness ratings from 35.88 to 94.38 and dividing the UV plot across realism levels shows that faces of the first level (fully realistic faces) replicate a curve like the one observed in Experiment 1 (*Figure 2.8).* Thus, the data suggest that the range of stimuli used in Experiment 1 corresponds to the rightmost part of the UV curve.

Figure 2.8

Linear slopes showing the relation between uncanniness and human likeness across faces' realism levels. The scatterplot is identical to the one in Figure 2.7, with the addition of depicting distortion levels.



Predictors of uncanniness. To test the effects of face realism, familiarity, orientation, and distortion on uncanniness, a linear mixed model was conducted with these predictors as fixed effects and base faces and participants as random effects. Results show significant main effects of realism ($t(1240) = 7.069$, $p < 0.001$), orientation ($t(11560 = -3.048$, $p = 0.002$), familiarity ($t(44470) = 2.512$, $p = 0.016$), and distortion ($t(10670) = 8.989$, $p < 0.001$). Furthermore, significant interactions were found between realism and familiarity

($t(11240)$, $p = -2.514$, $p = 0.012$), realism and distortion ($t(11560) = -4.494$, $p < 0.001$), orientation and distortion ($t(11560) = 4.667$, $p < 0.001$), and finally realism, orientation, and distortion ($t(11560) = -2.304$, $p = 0.0212$). No other term was significant ($R^2_{corr} = 0.565$).

In the next sections, results will be analyzed specific to variants of human famous faces and cartoon faces. Data for famous faces (realism levels 1 to 3) are summarized in *Figure 2.9*, and data for famous cartoon character faces (realism levels 4 and 5) are summarized in *Figure 2.10*.

Figure 2.9

Averaged uncanniness ratings across difference distortion levels (0 = base face), realism levels, and face conditions. Error bars represent standard errors.

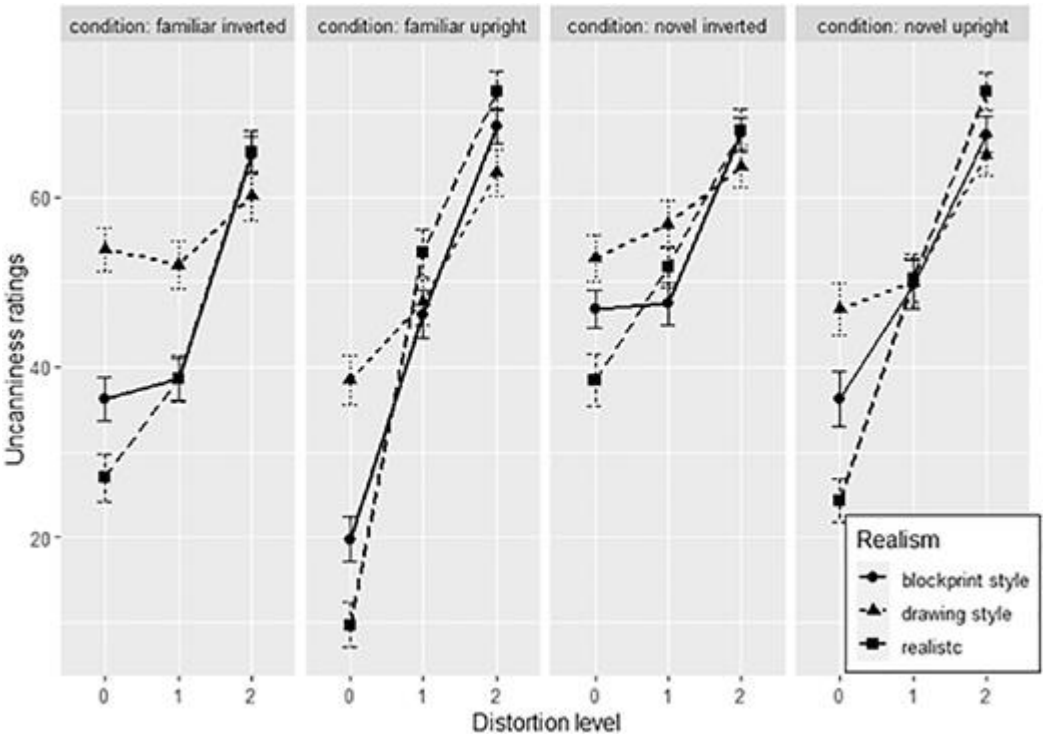
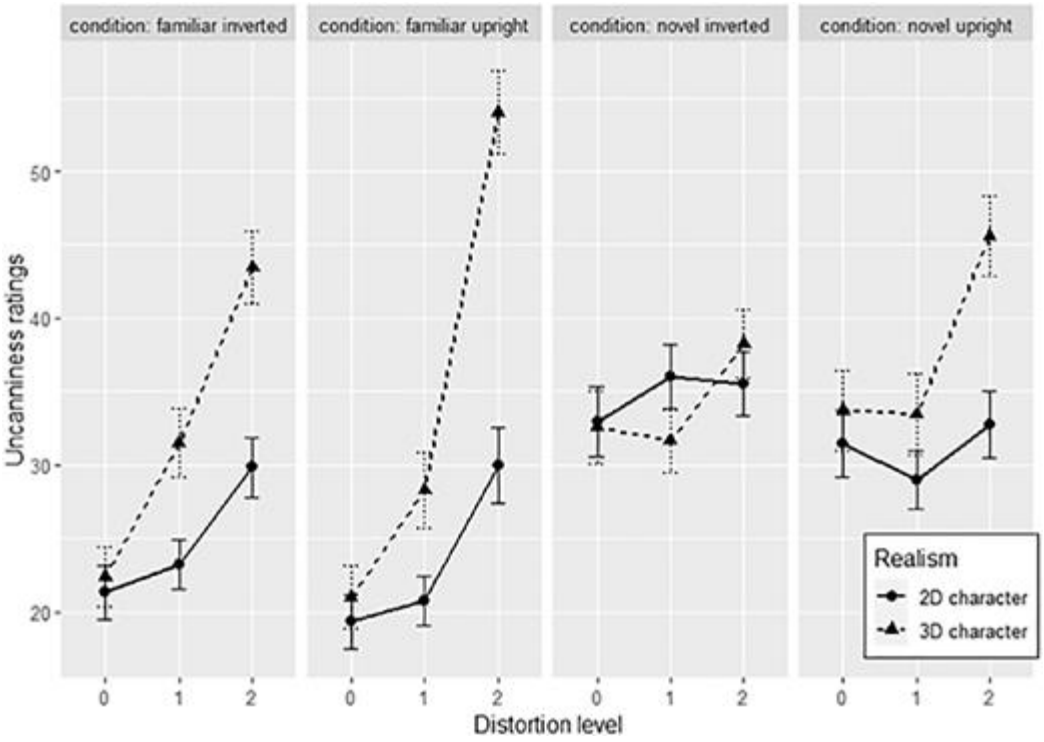


Figure 2.10

Averaged uncanniness ratings across difference distortion levels (0 = base face), realism levels, and face conditions. Error bars represent standard errors.



Post hoc Tukey tests were conducted to test the increase of uncanniness between distortion levels 0 to 1 and 1 to 2 for face realism levels and face conditions. For familiar upright faces, distortion significantly increased uncanniness across both distortion levels for fully realistic faces ($t(2049) = -8.545$, $p_{adj} < 0.001$ for the 0–1 distortion level; $t(2049) = -3.451$, $p_{adj} = 0.007$ for the 1–2 distortion level) and block print style faces ($t(2049) = -5.099$, $p_{adj} < 0.001$ for the 0–1 distortion level; $t(2049) = -4.732$, $p_{adj} < 0.001$ for the 1–2 distortion difference level), whereas for drawing-style faces, only the 1 to 2 distortion level difference significantly increased uncanniness ($t(2049) = -3.331$, $p_{adj} = 0.011$). Thus, for familiar upright faces, distortions increased uncanniness except for slight deviations in highly unrealistic faces.

For familiar inverted faces, all distortions of fully realistic faces increased uncanniness ($t(2049) = -3.332$, $p_{adj} = 0.011$ for the 0–1 distortion level; $t(2049) = -4.862$, $p_{adj} < 0.001$ for the 1–2 distortion level), but only the 1 to 2 distortion level in block print style faces ($t(2049)$

= −6.061, p_{adj} < 0.001) and none in the drawing-style inverted familiar faces. Thus, the uncanniness sensitivity for familiar inverted faces is decreased when faces are unrealistic.

For novel upright faces, again, uncanniness increased for realistic faces (t(2049) = −5.339, p_{adj} < 0.001 for the 0–1 distortion level; t(2049) = −4.191, p_{adj} < 0.001 for the 1–2 distortion level), but only at the 1 to 2 distortion level for block print style (t(2049) = −3.075, p_{adj} =0.025) and drawing-style (t(2049) = −3.132, p_{adj} = 0.021) faces. Thus, the uncanniness sensitivity for novel faces is decreased for slight deviations if faces are not realistic.

Finally, for novel inverted faces, only the 1 to 2 distortions increase uncanniness for realistic (t(2049) = −3.981, p_{adj} < 0.001) and block print style faces (t(2049) = −4.820, p_{adj} < 0.001). Thus, for novel inverted faces, only strong distortions increase the uncanniness in faces that are either realistic or slightly stylized.

In general, the results show that both familiarity, upright orientation, and high face realism increases the sensitivity of uncanniness to facial distortion. Thus, hypothesis 2 is supported.

Furthermore, post hoc Tukey tests were conducted to test the increase of uncanniness across distortion levels 0 to 1, 1 to 2, and 0 to 2 for 3D and 2D cartoon character faces, again across face conditions. For familiar upright faces, uncanniness increased only at the 1 to 2 distortion level for 3D faces (t(2122) = −7.723, p_{adj} < 0.001) and at the 0 to 2 distortion level for 3D (t(2122) = −7.339, p_{adj} < 0.001) and 2D faces (t(2122) = −2.987, p_{adj} = 0.034). Thus, strong deviations were uncanny in both 3D and 2D familiar upright faces.

For familiar inverted faces, uncanniness increased only at higher levels for 3D faces (t(2122) = −3.73, p_{adj} = 0.002 for the 1 to 2 distortion level; t(2122) = −5.803, p_{adj} < 0.001 for the 0 to 2 distortion level). Thus, only stronger deviations in more realistic familiar inverted faces increased uncanniness.

For novel upright faces, only 3D base faces increased in uncanniness at the 0 to 2 distortion levels ($t(2122) = -3.218$, $p_{adj} = 0.016$), suggesting that the sensitivity for uncanniness in novel upright faces is only present for strong deviations in more realistic faces.

Finally, uncanniness did not increase across distortion levels on any novel inverted faces. Thus, both novelty and inversion increase the range of acceptable face variation to the point where even strong deviations do not increase uncanniness.

Again, familiarity, upright orientation, and higher face realism increase the sensitivity of uncanniness to facial distortions. Thus, the results support hypothesis 2 even when using "natural" unrealistic base faces.

Moderating effect of distortion on base human likeness. Although a quadratic function akin to a UV plot can describe the data, the distribution based on realism level seen in *Figure 2.10* indicates that cartoonish characters can reach levels of human likeness akin to uncanny (deviating yet realistic) stimuli, despite not being uncanny themselves. Whereas the presence of non-uncanny stimuli at the same level of human likeness as uncanny stimuli can be observed in plots in previous research (e.g. Mathur & Reichling, 2016; Pütten & Krämer, 2014), it nevertheless begs the question whether the data can be explained by a function other than a polynomial plot, for example, as indicated in *Figure 2.10*, a moderated linear function: If the human likeness of a base face stimulus can be used as a proxy for the closeness to a typical face, it is expected that a higher degree of human likeness also activates a higher degree of configural processing and thus distortion sensitivity, which should reflect in a greater increase of uncanniness across distortion levels for more humanlike base stimuli. Specialized processing would be less important for judging deviations from less humanlike entities (e.g. cartoon faces), and thus deviations would be less increasingly uncanny. Higher face realism would then increase the slope of the effect of distortion on uncanniness, creating

a valley-shaped function when plotted across the data. Such a moderating linear function could underlie the relationship between human likeness and uncanniness typically observed as an UV plot, and has thus been investigated in the following exploratory analysis.

A linear mixed model with human li”enes’, realism, and distortion level as fixed factors, and participant and base face as random factors, was conducted. An interaction between human likeness, realism, and distortion could significantly explain the data ($t(11520) = -6.185$, $p < 0.001$, $R^2_{adj} = 0.55$). This interaction model was significantly better at explaining the data that the initial quadratic model ($\chi^2 = 968.74$, $p < 0.001$). Thus, a linear interaction model between a face’s realism level and distortion level across human likeness can better explain uncanniness than the typical quadratic function of human likeness.

Discussion

Uncanny valley and face distortion. The results show that a U-shaped, quadratic function best explained the data, analogous to a U-shaped valley found in previous uncanny valley research (see Diel et al., 2022). Although the UV has been associated with face distortions in past research (Diel & MacDorman, 2021; MacDorman et al., 2009; Mäkäräinen et al., 2014), this study is the first to properly locate face distortions on a UV curve. Results show that distorted version of real faces or stylized variants are located within the UV, compared to undistorted variants to the right and cartoon character faces to the left. Furthermore, the UV observed in this study could be divided into the pre-valley of cartoon faces and valley and post-valley consisting of real face variants, suggesting that an uncanny valley function consists of unrealistic, distant entities (e.g. cartoonish or exaggerated characters, or mechanical and stylized robots) left to the valley, imperfect or distorted variants of realistic human entities at the bottom of the valley, finally followed by fully human entities to the right of the valley. The higher sensitivity for configurations of realistic faces would then explain a harsher

judgment toward realistic entities failing to approximate the norm, compared to cartoonish or stylized unrealistic faces.

Face realism, familiarity, and orientation. The results show how face realism, familiarity, and orientation interaction with distortion levels to influence uncanniness ratings. Specifically, familiarity and upright orientation increase the sensitivity of uncanniness to facial deviations, which are again more sensitive for more realistic faces. Whereas even subtle deviations could increase the uncanniness in real faces, especially when they were upright and familiar, stronger deviations were needed to increase the uncanniness for stylized faces. Similarly, 2D cartoon faces had a wider range of acceptable, non-uncanny variations than 3D cartoon faces, and for the former, strong distortions only increased the uncanniness when faces were familiar and upright. The results thus indicate that a lower degree of realism generally increases the leeway of face variation, allowing the design of exaggerated facial proportions and expression without risking uncanniness (see also Green et al., 2008; MacDorman et al., 2009; Mäkäräinen et al., 2014). However, familiarity with a cartoon character further narrows the range of acceptable variations, potentially because a deviating familiar face is compared against the much narrower acceptable range of the familiar face representation rather than the acceptable range of all potential facial proportions. Similarly, inversion increases the range of acceptable variations, possibly by decreasing the ability to accurately process subtler configural information and thus potential deviations.

General discussion

Face familiarity and upright orientation increases distortion and uncanniness sensitivity. Whereas Mori's (2012) original graph is a good metaphor for possible negative reactions towards artificial humanlike entities, it does not capture some findings in UV research. First, sensitivity of the UV effect towards facial distortions is stronger for more realistic faces compared to less realistic faces (Green et al., 2008; Mäkäräinen et al., 2014). Second, a UV

effect has been observed with animal stimuli (e.g., Löffler et al., 2020; Mitchell et al., & MacDorman, 2011 Schwind et al., 2018; Yamada et al., 2013). Third, distortions of the structure of human faces elicits stronger uncanniness ratings than comparable distortions of the structure of cat faces or houses (Diel & MacDorman, 2021). Fourth, as observed in the present study, the sensitivity of uncanniness ratings for distortions is higher for familiar and upright faces compared to novel and inverted faces. Humans usually show a higher level of expertise and special processing for human compared to animal faces (Symons & Roberts, 2006). Furthermore, as face-typical processing is decreased for less realistic avatar faces compared to normal faces (Kätsyri, 2018), a higher level of perceptual experience with a category of faces may increase sensitivity to deviation. Thus, a mechanism underlying the UV effect may be the enhanced ability to detect deviations from familiarized objects and categories, possibly due to an increased experience with recognizing and differentiating individual exemplars. This model would also predict an uncanny valley prevalently for closely human entities with weaker variants for other stimuli like animals and familiar objects. Last, as a topic for future research, manipulating perceptual expertise for a stimulus category should increase the distortion sensitivity of uncanniness ratings.

While Chapter 2 found that markers of specialization predicted uncanniness sensitivity, it relied on correlates of specialization (familiarity, orientation, realism). A direct statistical link between a measure of specialization (e.g., the inversion effect) and uncanniness sensitivity is not yet established. The following chapter will fill this gap.

Chapter 3: Smoothing the uncanny valley: Specialization moderates the linear effect of deviation on uncanniness

Methods, experiment, and large portions of the introduction and discussion in this chapter is currently in review in the journal *Computers in Human Behavior*.

Introduction

Even though the uncanny valley effect is a well-replicated phenomenon (see Diel et al., 2022), conceptual arguments (e.g., the relevance of the human likeness axis; see below) and a lack of parsimony (complex cubic functions are unusual in nature) begs the question of the validity of its first theoretical proposal as a cubic relationship between human likeness and likability or related ratings (here called the *initial uncanny valley model*). Here, it is investigated and discussed whether this initial uncanny valley model can instead be rethought as a moderated linear function between typicality/deviation, likability/uncanniness, and the degree of specialization. Such a redesigned model of the uncanny valley is capable of explaining a broader range of observations beyond the initial uncanny valley model while not suffering from its disadvantages. Hence, a direct statistical link between specialization, deviation, and uncanniness is investigated.

The statistical value of cubic relationships

Although simplicity is preferred in scientific explanations, interactions between variables do not always follow the simplest, linear relationships: For example, psychological models often follow quadratic functions, such as stress models or the Yerkes-Dodson law (Yerkes & Dodson, 1908). Polynomial degree reduction can help to simplify otherwise complex statistical patterns into simple laws. For example, quadratic relationships can be reduced to "deviation-from-optimum" relationships for which changes from an optimal value of one variable (e.g., pressure) linearly change the value of the other variable (e.g., performance). The non-monotonic nature of the uncanny valley requires a model that is cubic in nature with

a part that is concave up and a part that is concave down. Very few phenomena in nature follow a cubic function like this and so, if the uncanny valley is describing the simple relationship between two properties then its relationship would be fairly unique. An attractive, simple alternative to a cubic model is to introduce a third variable that moderates the effect of the independent variable on the dependent variable depending on the degree of the former. In terms of the uncanny valley, a high level of human likeness may sensitize the effect of deviation from typical appearance on likability (or uncanniness), increasing the uncanniness if a stimulus is anomalous. At lower levels of human likeness meanwhile, the effect of typicality on likability would be less pronounced. When reduced to only two variables and with particular selection of exemplars, a simple moderated linear function would take the form of a complex cubic model akin to the uncanny valley.

"Human likeness" and the uncanny valley

The initial uncanny valley model focused on a human likeness dimension (Mori, 2012) and remains an essential part of the uncanny valley's understanding today (Diel et al., 2022; Mara et al., 2020; Zhang et al., 2020). The focus on human likeness may be due to the uncanny valley's relevance in robotics or creation of virtual entities with humanlike appearance. However, as mentioned previously, uncanny valley effects have also been observed using animal stimuli (Diel & MacDorman, 2021; Löffler et al., 2020; MacDorman & Chattopadhyay, 2016; Rativa et al., 2022; Schwind et al., 2018; Yamada et al., 2013). Yet when including both human and non-human animal stimuli in one dataset, focusing only on a human likeness dimension would not sufficiently represent the animal-related uncanny valley effects (namely, that animal stimulus manipulations can increase uncanniness). Furthermore, uncanny valley-like effects have been observed for inanimate categories like written text or physical places (Diel & MacDorman, 2021; see Chapters 5 and 6). A two-variable model

including only likability/uncanniness and human likeness would insufficiently account for uncanny valley effects beyond human stimuli.

Typicality/deviation and likability/uncanniness

Given the range of stimulus types for which uncanny effects have been observed, a human likeness dimension seems insufficient. A more general approach is related to changes in likability (or uncanniness) dependent on a stimulus' typicality (or degree of deviation): deviating stimuli or patterns tend to be disliked across categories, and individual differences in the degree of aversion can be transferred across stimulus categories (Gollwitzer et al., 2017). Aversion caused by deviation may be caused by increased processing disfluency (Winkielman et al., 2003) or violations of expectations in predictive coding (Friston, 2010). Such mechanisms would not be bound to a human likeness dimension and could explain uncanny valley effects across animal (e.g., Schwind et al., 2018) and inanimate object (e.g., see Chapters 5 and 6) categories.

However, the effect of typicality (or deviation) on likability (or uncanniness) is not consistent across categories: For example, analogous distortions increase uncanniness more in human compared to cat faces, and cat faces compared to buildings (Diel & MacDorman, 2021). Furthermore, effects of facial distortion on likability are more pronounced in more realistic faces (Chapter 2; Green et al., 2008; MacDorman et al., 2009; Mäkäräinen et al., 2014). Effects of deviation on uncanniness thus seem more pronounced in some categories (e.g., realistic and human faces) than others (e.g., unrealistic and animal faces). While the initial uncanny valley does not provide a clear solution on this, a redefined model of the uncanny valley may benefit from adding a moderating variable defining the strength of effect of deviation on

Specialization as a moderator variable

A high sensitivity to deviations in especially realistic human-related stimuli (e.g., faces) may be due to a high degree of processing specialization for such categories. Humans are highly specialized for upright human faces (Gauthier & Nelson, 2001; Maurer & Werker, 2014; Rhodes, Brake, Taylor, & Tan, 1989), which enables assessment of facial identity and aesthetics based on feature-relational information; a process that is disturbed when faces are presented inverted (Carbon & Leder, 2006; Mondloch et al., 2002). A higher degree of specialization enables a more detailed aesthetics assessment by a higher sensitivity to slight differences in facial structure (Chapters 2). Perceptual specialization is not exclusive to faces (Gauthier & Nelson, 2001) and trained specialization for an otherwise novel category increases the uncanniness of distorted variants compared to non-distorted variants, and compared to distorted variants without such training (see Chapter 4). As specialization is high in human stimulus categories (e.g., faces, bodies, voices, motion), deviations in these categories would appear especially uncanny, leading to the uncanny valley effect at high levels of human likeness.

The reduced ability to recognize a face when inverted (compared to upright) has been used as a measure of a degree of specialization. Face inversion effects are reduced for less realistic faces (e.g., computer-generated, virtual) compared to typical human faces (Balas & Pacella, 2015; Crookes et al., 2015; Di Natale, Simonetti, La Rocca, & Bricolo, 2023). Higher specialization for more realistic faces could explain a higher sensitivity to deviations described above (Chapter 2; Green et al., 2008; MacDorman et al., 2009; Mäkäräinen et al., 2014). Furthermore, as specialization is less pronounced in less realistic entities (including mechanical robots; Sacino et al., 2022; Zlotowski & Bartneck, 2013), tolerance for atypicalities or deviations should be higher in these categories, leading to a lower likelihood of the occurrence of uncanny exemplars. Together with a high deviation sensitivity in more

realistic humanlike stimuli, the uncanny valley effect would thus emerge across the dimension of human likeness (see *Figure 1.2*).

Thus, the degree of specialization is a suitable third variable candidate for simplifying the uncanny valley as a moderated linear function. In faces, specialization can be quantified using the face inversion effect (difference of recognition abilities for upright compared to inverted faces). Furthermore, a moderated linear relationship is not bound to the human likeness dimension and can be applied to animal or inanimate stimuli, a limitation of the initial uncanny valley model. In summary, a moderated linear function of specialization, typicality/deviation, and likability/uncanniness may explain a wider range of data with higher accuracy than a nonlinear initial uncanny valley model (Mori, 2012).

Experiment 3

The aim of this work is to investigate whether the uncanny valley can be better understood as a moderated linear function of deviation, uncanniness, and specialization. Specifically, it is investigated whether specialization in different face types (face inversion effect) moderates the effect of face distortion (incremental changes in face feature positions) on uncanniness: Uncanniness is expected to increase with facial distortions, and this effect should furthermore increase with specialization in the face group.

Research question and hypotheses

First, the face inversion effect is replicated for each face category, and it is replicated whether the effect is stronger for more realistic compared to less realistic faces (e.g., Crookes et al., 2015; Sacino et al., 2022):

1. A face inversion effect is stronger for more realistic human faces (human and cartoon faces) compared to less realistic faces (drawing and robot faces) (*inversion effect hypothesis*)

Second, the conventional uncanny valley is investigated by testing whether a polynomial (quadratic or cubic) function of human likeness ratings can explain uncanniness ratings (Mori, 2012):

2. A polynomial function of human likeness can explain the uncanniness across faces better than a linear function (*uncanny valley hypothesis*)

Third, it is suggested that a moderated linear function underlies the uncanny valley: a higher sensitivity to distortions is proposed for more humanlike or realistic faces, which would increase the relative uncanniness caused by deviations (Chapter 2). As specialization sensitizes the detection of changes and distortions (see Chapter 4), it is tested whether by-participant recognition accuracy differences between upright and inverted faces for each condition (face inversion effect as a marker of expertise) predict the effect of distortion on uncanniness:

3. Degree of Inversion effect predict uncanniness caused by distortion (*moderation hypothesis 1*)

Finally, to investigate whether a moderated linear function it is tested whether specialization as a moderator for distortion and uncanniness can explain the data better than a polynomial function of uncanniness and human likeness:

4. A moderated linear function of face distortion level, uncanniness, and inversion effect can explain the data better than a nonlinear function of uncanniness and human likeness (*moderation hypothesis 2*)

Methods

Participants

As previous research found odds ratio (OR) values of 0.62 (converted to a Cohen's d = 0.264), for inversion effects in robot stimuli (Sacino et al., 2022), a power analysis with an effect size of d = 0.264 revealed that 120 participants is enough for a power of $1 - \beta = 0.8$. Participants (M_{age} = 19.4, SD_{age} = 0.84) were 120 undergraduate Psychology students of Cardiff University; 103 identified as female and 17 as male.

Material

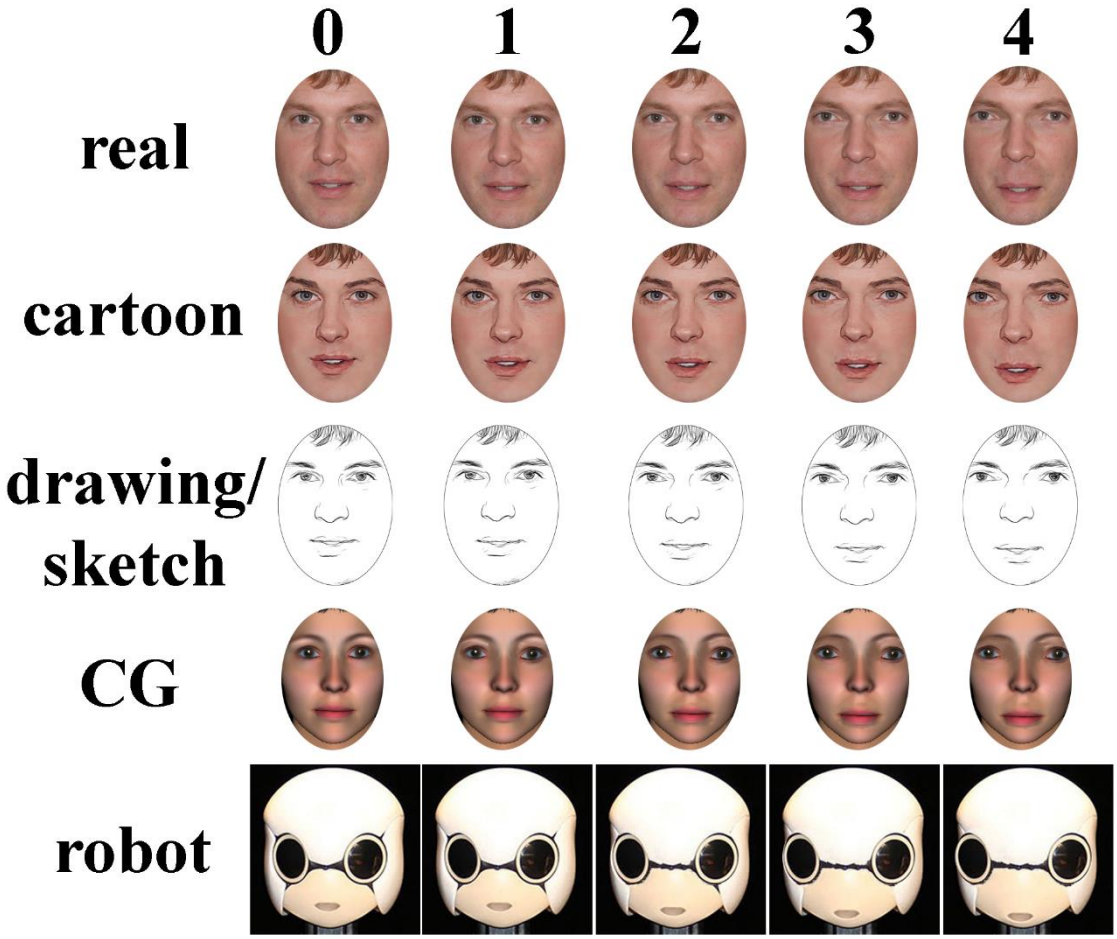
Human faces were selected from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015). Different sets of 12 faces (three female, three male) were used for each of the first three levels of realism (real, cartoon, drawing). Cartoon and drawing faces were created using the cartoon character and sketch character tools of VanceAI toongineer (https://vanceai.com/toongineer-cartoonizer/). Realism level 4 (CG) faces were created using FACSGen. Finally, realism level 5 (robot) faces were selected from a previous study locating a wide range of robot faces before the uncanny valley (Mathur & Reichling, 2016).

All faces were distorted in the same manner: Distance between eyes were incrementally increased in five faces per group, and decreased in the other five faces, by 10% of the eyes' horizontal length. In addition, the position of the mouth was either incrementally increased or decreased (each in five faces per group) by 25% of the mouth's vertical length. A total of five distortion levels, including the original, were created. Distortions were manipulated in both directions (e.g., eye distances were either increased or decreased) to control for different types of facial distortions.

Finally, half of the undistorted base faces of each face realism group were inverted for the face recognition task. Stimuli divided by condition can be seen in *Figure 3.1*.

Figure 3.1

Upright stimuli divided by distortion (horizontal axis; 0 to 4) and face type (vertical axis; real, cartoon, drawing, CG, robot) conditions. Note: Faces were also presented inverted real, cartoon, and drawing faces depicted here were not used in the experiment. The faces were artificially created by the StyleGAN generative network (Karras, Laine, Aittala, Hellsten, Lehrinen, & Aila, 2020).



Procedure

Face recognition task. The face recognition task consisted of an encoding part and a recognition part. Only undistorted faces were used in the face recognition task. In the encoding part, participants viewed a total of 60 faces (12 faces per realism level; half upright,

half inverted; half female, half male) sequentially in a random order. Participants were allowed to view each face for as long as they wanted. In the recognition face, an additional novel 60 faces (10 faces per realism level; half upright, half inverted) were shown to the participants together with the learnt faces, and participants were asked to indicate for each face whether they have seen the face in the encoding phase. Again, participants had an indefinite amount of time to decide for each face while simultaneously viewing the face.

Face rating task. Both undistorted and distorted faces were used in the rating task. In the face rating task, each face was shown the participants in a random order, together with three scales: *uncanny/eerie*, *strange/weird*, and *realistic/humanlike*. Scales were shown in the same order as mentioned here, and the *strange/weird* scales were reversed. Participants were to rate each face on each scale ranging from 0 to 100. Each face was shown for the entire time until participants responded for each scale. A total of 300 faces (10 faces per 6 realism level and 5 distortion levels) were rated. Participants had an indefinite amount of time to rate each face while the face was presented.

Data analysis and availability

RStudio and JASP were used for data analysis. The degree of expertise was calculated by using the differences of means and standard deviations between upright and inverted faces, and applied for each face realism group.

Ethics statement

Research was conducted in accordance with the Declaration of Helsinki. The study was approved by the Cardiff University ethics committee board (EC.23.01.10.6716).

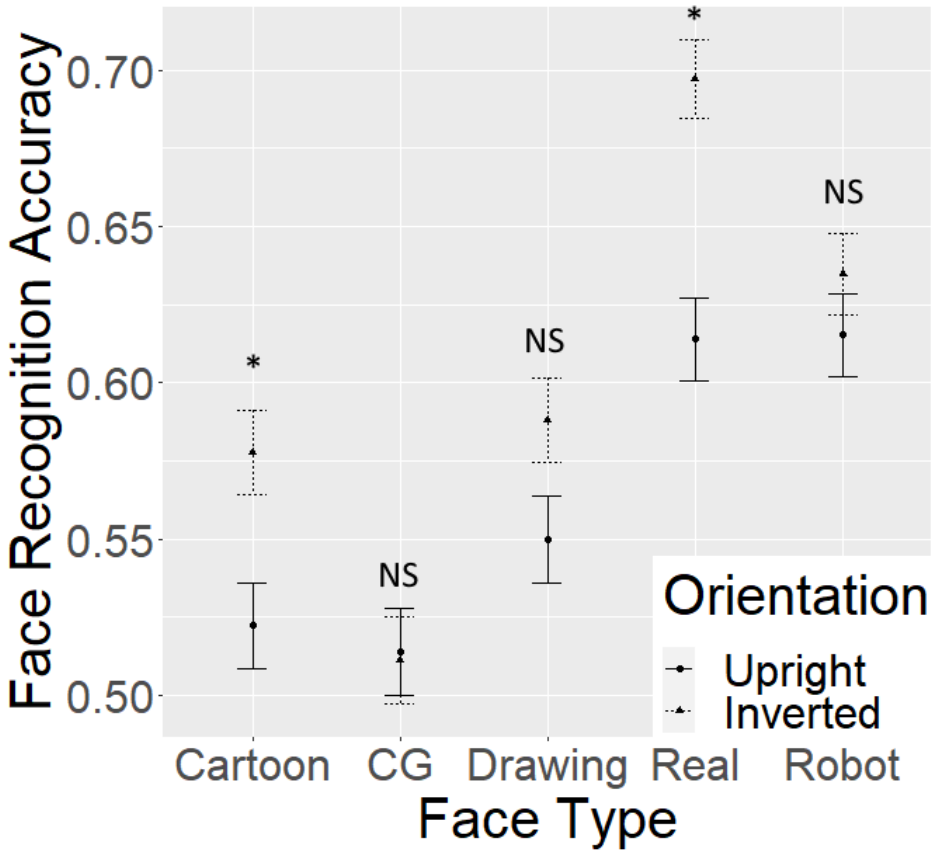
Results

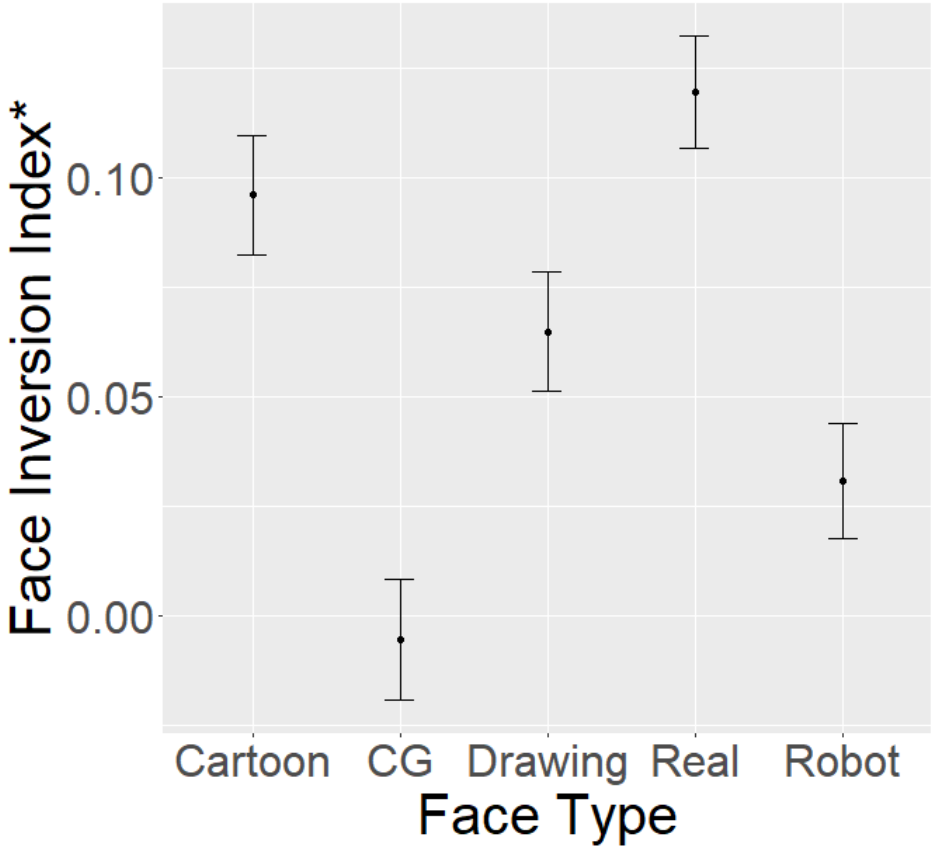
Face inversion effect

Recognition accuracy was calculated by averaging by-participant numbers of correct responses participant for each face condition. The face inversion effect was then tested by calculating the differences between upright and inverted face recognition accuracy for each face condition. The data is summarized in *Figure 3.2*.

Figure 3.2

Average face recognition accuracy across face type and condition (A) and level of face inversion effect (difference between upright and inverted face recognition accuracy) across face type (B). For A, asterisks show significantly higher recognition rates for upright compared to inverted faces while "NS" indicate no significant increases. Error bars indicate standard errors.





*1 - (upright recognition accuracy / inverted recognition accuracy)

A within-subject and within-base stimulus ANOVA on face recognition accuracy with face type and orientation as factors found significant main effects of face type ($F(1,118) = 28.57$, $p < .001$, $\eta^2_p = .001$) and orientation ($F(4,115) = 37.383$, $p < .001$, $\eta^2_p = .01$) and a significant interaction ($F(4,115) = 3.32$, $p = .01$, $\eta^2_p = .001$). Results are indicative of a face inversion effect that differs between face types.

To investigate the face inversion effect per face type, post-hoc comparisons with Bonferroni-adjusted *p*-values between upright and inverted faces were performed for each face type. Upright faces were significantly better recognized than inverted faces for real faces ($t(14271) = 4.6$, $p_{adj} < .001$, $d = 0.49$) and cartoon faces ($t(14271) = 3.07$, $p_{adj} = .006$, $d = 0.32$), but not for face drawings ($t(14271) = 2.11$, $p_{adj} = .09$), CG faces ($t(14271) = 0.16$, $p_{adj} = 1$), or robot faces ($t(14271) = 1.07$, $p_{adj} = 1$). Thus, inversion effects were observed for real and cartoon faces, but not for the other face conditions. Thus, hypothesis 1 (*face inversion hypothesis*) is supported.

Uncanny valley

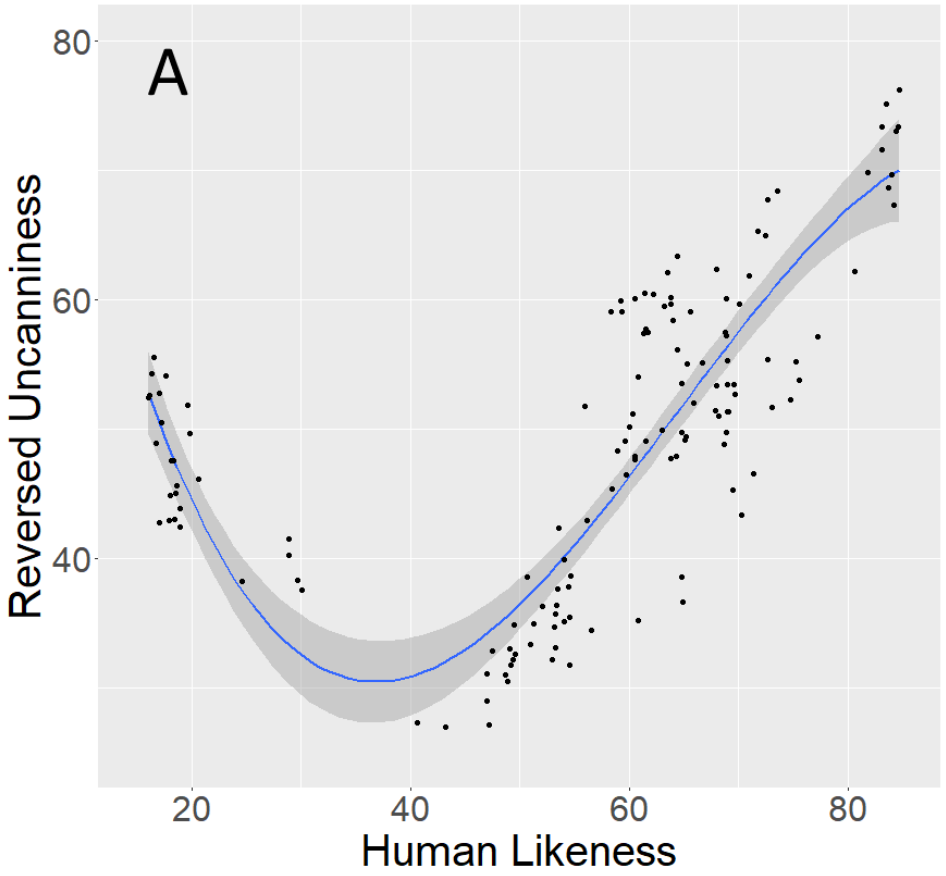
The uncanny valley hypothesis was investigated by testing whether a polynomial relationship between human likeness and uncanniness can explain the data better than a linear function. Linear mixed models with participants and base faces as random effects and linear, quadratic, and cubic function of human likeness as fixed effects were performed as predictors of uncanniness. Significant linear ($t(1561) = -5.5$, $p < .001$), quadratic ($t(1561) = 9.29$, $p < .001$), and cubic ($t(1561) = -55.42$ $p < .001$) functions of human likeness were found. Furthermore, the quadratic (AIC = 138004)[2] model was a significantly better fit compared to the linear (AIC = 138897) model ($\chi^2 = 894.58$, $p < .001$), while the cubic (AIC = 137770) model was a

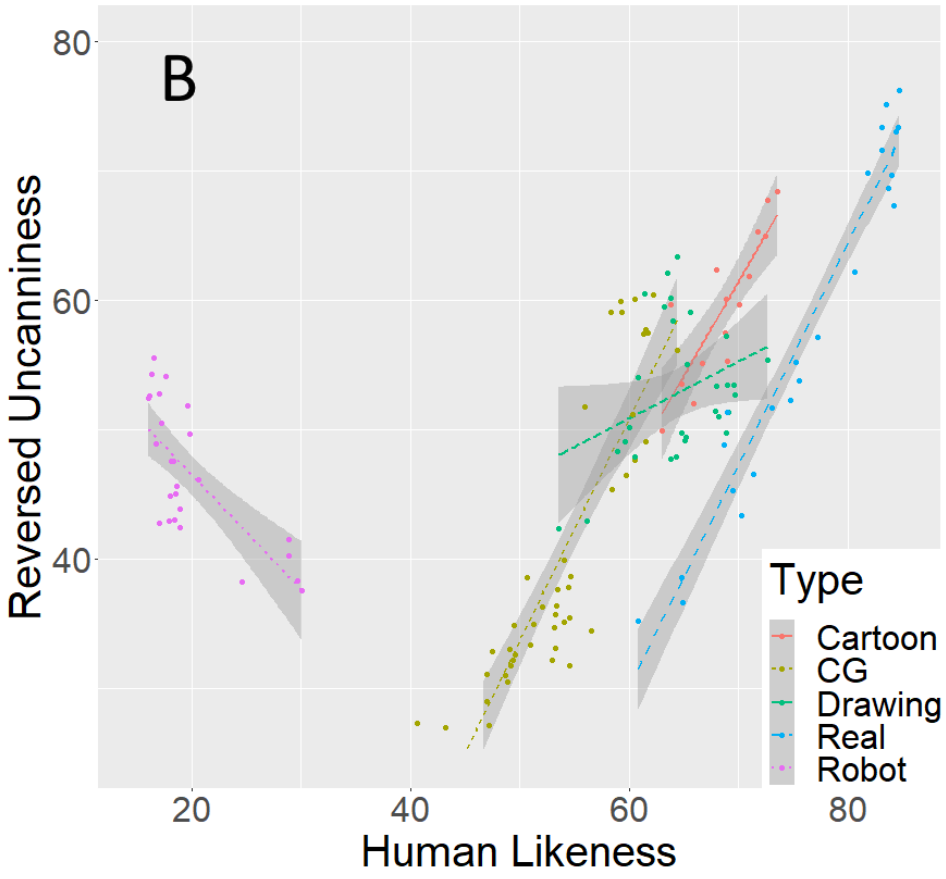
[2] AIC (Akaike information criterion) is a measure of how well a model fits the data. Higher values correspond to stronger deviations of the data from the model. Hence, models with lower AIC values are considered a better fit.

better fit than the quadratic one (χ^2 = 235.96, p < .001). Thus, the cubic model of human likeness (R^2_c = .43) could best explain uncanniness. Thus, hypothesis 2 (*uncanny valley hypothesis*) is supported. The fit is depicted in *Figure 3.3A*.

Figure 3.3

Nonlinear (A) and moderated linear (B) fits on uncanniness. Gray areas represent standard errors. Dots show individual stimulus values averaged across participants, and are the same for both A and B. Figure 3.3A depicts a nonlinear function of human likeness across all face types while Figure 3.3B depicts linear relationships between human likeness and uncanniness for each base stimulus, categorized by types.


A
Reversed Uncanniness
80
60
40
20
40
60
80
Human Likeness

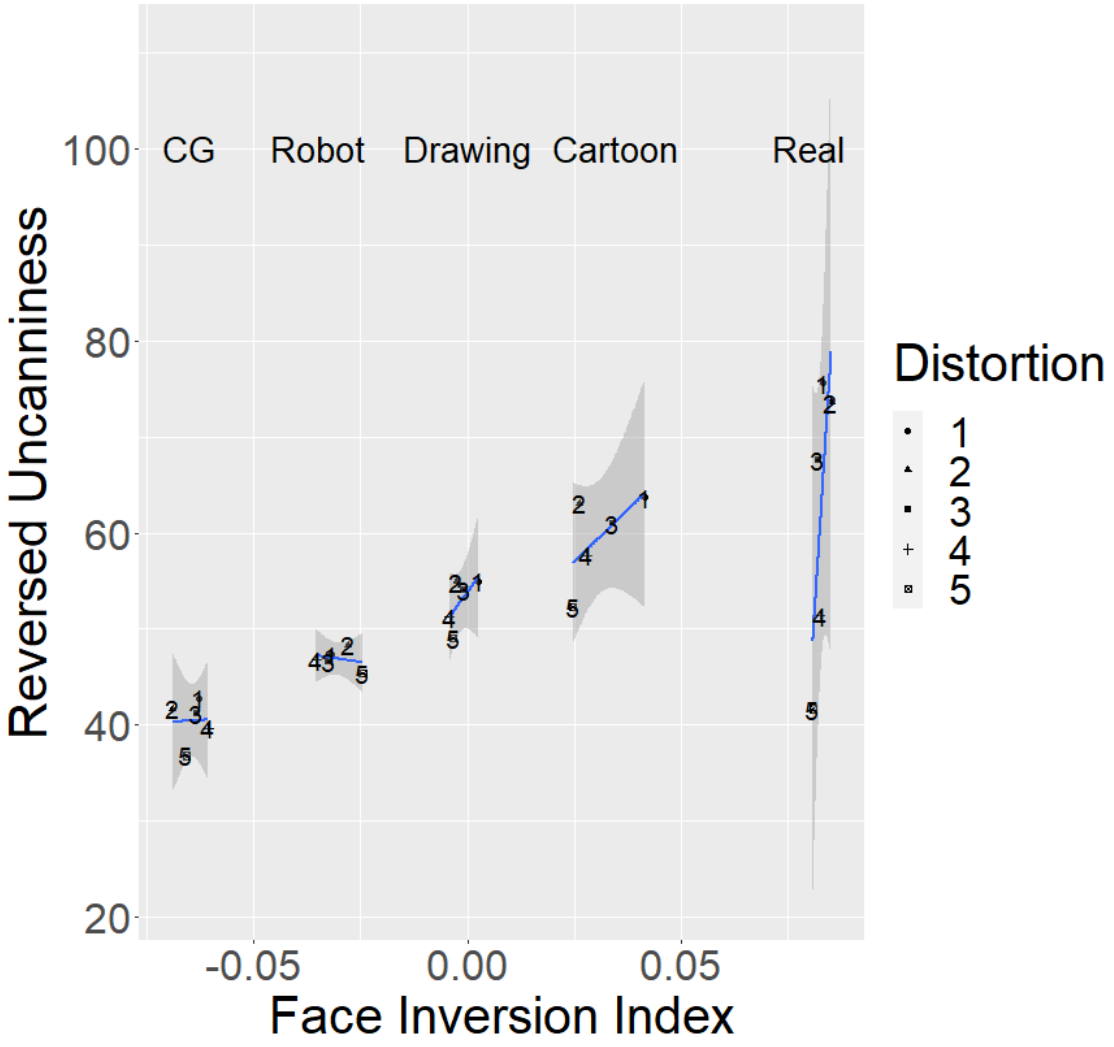

B
Reversed Uncanniness
80
60
40
20
40
60
80
Human Likeness
Type
Cartoon
CG
Drawing
Real
Robot

A moderated linear function

Hypothesis 3 predicts that a moderated linear function of distortion and expertise (face inversion index) explains uncanniness, and better than a polynomial function. Face inversion index as a proxy to expertise has been calculated as 1 minus inversion recognition rate divided by upright recognition rate. Linear mixed models with participant and base face as random factors and distortion and face inversion index as random effects found significant main effects of distortion ($t(15496) = 2.83$, $p < .001$) and face inversion index ($t(15559) = -13.41$, $p < .001$), as well as a significant interaction ($t(15496) = 3.57$, $p < .001$; $R^2_c = .37$). The interaction is summarized in *Figure 3.4*: face types are divided based on their average IEI (x-axis), and uncanniness levels are plotted for each stimulus distortion level. Linear slopes depict the average change in uncanniness across distortion levels, which are steeper for stimulus categories with a higher IEI.

Figure 3.4

Moderated linear function of expertise and uncanniness. Groups are divided by actor type. Gray areas indicate standard errors, and numbers indicate distortion levels averaged across stimuli and participants.



In summary, with a higher degree of a face's inversion effect, the effect of distortion on uncanniness increased. Thus, hypothesis 3 (*moderation hypothesis 1*) was supported

Finally, to test whether a moderated linear function can best explain uncanniness, a linear mixed model with participants and base faces as random effects and actor type, distortion, face inversion index, and human likeness as fixed effects has been calculated and tested against the cubic function of human likeness. The moderated linear model (R^2_c = .46; AIC = 137213) could explain uncanniness better than the cubic (R^2_c = .43; AIC = 137770) model (χ^2 = 628.96, p < .001). Thus, hypothesis 4 (*moderation hypothesis 2*) was supported.

Discussion

Summary of results

Here it was investigated whether face inversion index (a marker for specialized processing) can moderate the sensitivity of uncanniness to facial distortions, and whether such a moderated linear relationship can explain the data better than a traditional polynomial uncanny valley plot (Mori, 2012).

In accordance with previous research, the strength of inversion effects differed across face conditions (e.g., Di Natale et al., 2023; Sacino et al., 2022): Specifically, as seen in *Figure 3.2*, inversion effects were found for real human faces and cartoon faces, but were not found for drawing-style faces, CG faces, and robot faces.

As shown in *Figure 3.3A*, a function of human likeness and uncanniness indicative of an uncanny valley was found. However, as seen in *Figure 3.3B*, the same data can also be plotted as a moderated linear function when data is divided by face type: the changes of uncanniness across human likeness was different across face type in a linear manner. In addition, these results are surprisingly similar to the predictions in the hypothetical reinterpretation of the uncanny valley as a moderated linear function depicted in *Figure 1.1*.

Since inversion effect scores differed between face types, and because specialization may sensitize the detection of distortions (Chapter 2), it stands to reason that specialization could moderate the uncanniness caused by distortion. As seen in *Figure 3.4*, higher face inversion scores indeed correlated with an increase in uncanniness across distortion levels.

In summary, it is here suggested that the uncanny valley effect can be rethought of as a moderated linear function: at lower human likeness levels, specialization is relatively low, leading to a low sensitivity to deviations (see the robot faces in *Figure 3.3*), and because an increase in humanlike appearance generally makes a character more appealing (Mara et al.,

2022), a wide variation of near humanlike designs remains acceptable despite a spectrum of exaggerated or distorted face or body configurations. With higher levels of human likeness specialization increases, leading to a higher sensitivity to even subtle distortions. When humanlike appearance is attempted in the design of realistic androids, slight deformations may be recognized through specialized processing that would be otherwise acceptable in less realistic entities, creating the uncanny valley (*Figure 3.3*).

Uncanny valley: A function of specialization, deviation, and uncanniness

The advantage of a rethought moderated uncanniness function lies in the range of explainable data: the traditional uncanny valley model was restricted to a dimension of human likeness (Mori, 2012), which already complicated interpretations of an uncanny valley modeled for animal stimuli (e.g., Schwind et al., 2018). In addition, the traditional uncanny valley model was inadequate to explain a stronger uncanny valley effect for human compared to animal stimuli (Diel & MacDorman, 2021; Diel et al., 2022). A moderated linear function meanwhile can generalize predictions onto any stimulus category with quantifiable levels of specialization: Not only can a higher sensitivity towards distorted human faces compared to animal faces be explained by a higher level of specialization for the former; the theory also encompasses uncanniness in inanimate categories like architecture (see Chapters 5 and 6). In addition, the moderated linear model can explain why the sensitivity to facial distortion is increased for more realistic faces (Chapter 2; Green et al., 2008; MacDorman et al., 2009; Mäkäräinen et al., 2014), increased sensitivity to distortions for familiar compared to unfamiliar faces (Chapter 2; Jung, Lee, & Choi, 2022), and a higher uncanny valley effect for own-ethnicity compared to other-ethnicity faces (Saneyoshi, Okubo, Suzuki, Oyama, & Laeng, 2022).

Although the moderated linear model can function as a descriptive statistical model, it provides no explanation of *why* a deviating stimulus may appear uncanny. Underlying

neurocognitive mechanisms compatible with the moderated linear model may provide a link between deviation and devaluation.

Processing disfluency predicts aesthetic devaluation of stimuli that deviate from (proto-)typical appearance (Winkielman et al., 2003). Faces that are disfluent due to categorical ambiguity are devaluated (Halberstadt & Winkielman, 2014), just as faces further from a distributional center of facial structure (Dotsch, Hassin, & Todorov, 2017). Deviating faces may thus be devalued due to their relative distance from typical face appearance. On that account, specialized processing may increase relative disfluency by activating more specific processing dimensions (e.g., configural structure) on which a stimulus may deviate. Thus, given the same physical distortions, a deviating in a stimulus may be processed more disfluent if it is processed in a specialized manner, compared to the same stimulus if it is not processed in a specialized manner (see Chapter 4).

The *Prediction error* model proposes that the mind generates predictive models of the world which are contrasted with perceptual information, and discrepancies between prediction and input (i.e., violations of expectations) elicit predictive errors (Friston, 2010). Categorical specialization may sensitize the predictions of e.g., facial structure, leading to a relatively stronger prediction error compared to a stimulus not belonging to a specialized category. As negative affect can be caused by prediction errors (Van de Cruys, 2017), a relatively high prediction error caused by a stimulus deviating from typical appearance of a specialized category may be evaluated especially negatively.

However, while both theories may predict general devaluation of deviating stimuli, their predictions lack the specific negative experience characteristic of the uncanny valley: *eerie*, *strange*, *weird*, or *uncanny* (Diel et al., 2022; Ho & MacDorman, 2017), and has been associated with both disgust and fear (Ho, MacDorman, & Pramono, 2008). However,

uncanniness experiences specifically were never associated with disfluency or prediction errors. In the context of predictive coding, prediction errors were even associated with positive aesthetic evaluations (e.g., Delplanque, De Loof, Janssens, & Verguts, 2019). Hence, exact associations between general, wide-ranging neurocognitive processes like predictive coding and specific negative experiences like uncanniness still remain unclear and may be a topic of future research.

Chapters 2 and 3 have established statistical links between a stimulus' level of specialization and sensitivity to distortion: Chapter 2 showed that markers correlating with specialization (e.g., familiarity, realism, orientation) correlate with a higher sensitivity to, and higher uncanniness ratings of, distortions. Chapter 3 presented research showing that a direct marker of specialization (face recognition inversion effect) could predict the uncanniness sensitivity of distortions. However, conclusions about a causal link between specialization and distortion sensitivity cannot yet be drawn. Chapter 4 aims to investigate such a causal link.

Chapter 4: The deviation-from-familiarity effect: Expertise increases uncanniness of deviating exemplars

Methods, experiment, and large portions of the introduction and discussion in this chapter has been published in the journal *PLOS ONE* (Diel & Lewis, 2022b).

Introduction

Although Chapters 2 and 3 provide evidence of the association between specialization and deviation sensitivity on uncanniness, a causal link has not yet been investigated.

A manipulation of specialization

Greebles are designed to be individually recognizable based on differences in single features that follow the same configural pattern (Gauthier & Tarr, 1997), and trained expertise in Greebles has been show to approximate behavioral and neural correlates of face processing (Gauthier & Nelson, 2011; Gauthier et al., 1998). Thus, greeble expertise training is a viable candidate method to investigate the effect of prolonged exposure on the uncanniness of configural distortion. Thus, the present research will focus on the cognitive causes of uncanniness as predicted by previous theories on the uncanny valley. An uncanny valley function is not replicated here, nor are humanlike stimuli used. Nevertheless, the results may provide important insights into the cognitive mechanisms underlying the uncanny valley phenomenon.

Experiment 5

The present work will investigate cognitive causes of experiences of uncanniness and abnormality. It is proposed that aesthetic judgment of exemplars based on their distance to a category's centre or prototype should depend on the degree of perceptual expertise with the category. Exemplars distant from the "normal range" of observed exemplars surrounding the centre should be rated as more uncanny, while stimuli closer to the centre should appear more

attractive. Thus, greeble expertise training should increase the attractiveness of averaged greebles relative to normal greebles, while increasing the relative uncanniness of configurally distorted greebles deviating from the norm.

Research question and hypotheses

The current study is the first to investigate the effect of expertise training on these two phenomena: The uncanniness of distorted category exemplars and the attractiveness of averaged (prototypical or blended) category exemplars. Participants rated *uncanniness* and *attractiveness* of normal, averaged, and distorted greebles either after 5-day expertise training (*training condition*) or without expertise training (*control condition*).

The following hypotheses were tested:

- Distorted greebles are rated as more *uncanny* after training (*training condition*) than distorted greebles without training (*control condition*) when compared to the normal greebles.
- Distorted greebles are rated as *less attractive* after training (*training condition*) than distorted greebles without training (*control condition*) when compared to the normal greebles.

Methods

Participants

Participants were 45 Cardiff University psychology students randomly split into 21 participants in a *training group* and 24 in a *control group*. Participants had a mean age of M_{age} = 19.52, SD_{age} = 1.42, and 36 were female. Because the interpretation of the results was predominantly based on Bayesian inference which is not affected by sample size, sample size was decided on the Bayesian stopping rule after collecting an initial set of participants, and because evidence either in favor or against the null hypothesis was already present with

the initial sample size, data collection was stopped at that point (Rouder, 2014; Wagenmakers et al., 2019). For *p*-value statistics, a post-hoc power analysis revealed that a power of 1- β = 0.8 would be achieved with the given sample size and an effect size of $d = 0.4$.

Stimuli

Greeble training set. A set of 30 asymmetrical greeble stimuli from the tarrlab stimulus database (see https://sites.google.com/andrew.cmu.edu/tarrlab/stimuli) were used for the study. The greeble set consisted of six individual greebles per five families. Greeble families differed by having distinct body shapes. Within a family, individual greebles shared a body shape but differed in the shape of their four features. Features' positions were approximately the same for all greebles. Each greeble was matched with an individual label (four letter neologisms starting with a consonant), and each family with a family label (four letter neologisms starting with a vowel). These were the greebles used for the training.

Greeble test set. In addition to 25 of the 30 training set greebles (five per family), the test set consisted of ten distorted variants and six morphed variants. Ten training set greebles (two individuals per family) were used to create configurally distorted variants by changing the position of three of the four of the greebles' attached body parts. Only the body parts' relative positions and angles were changed while the body parts themselves and the greeble bodies remained unedited, to create distortions on a configural level rather than on a featural level. All distorted greebles were edited in the same manner and the changes will be reported in degrees and percentages of the greebles' total size, and the changes are visualized in *Figure A1*: The upper right body part (P1 in *Figure A1*) was placed 30% downwards on the body and rotated by about 45 degrees upwards. The upper left body part (P2 in *Figure A1*) was mirrored on the vertical axis and placed about 10% to the right and 10% upward, around the centre of the greeble head. Finally, the leftmost body part (P3 in *Figure A1*) was positioned 10% to the right and 15% upwards towards the "neck" area of the body, and angled by 30

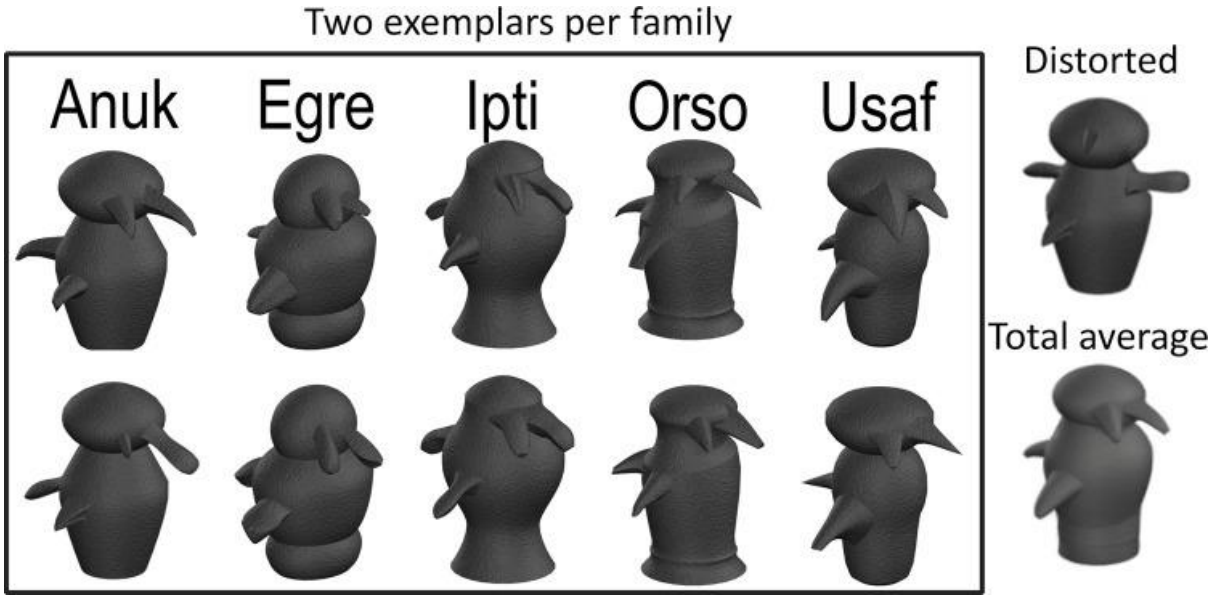
degrees upwards. *Figure A2* depicts examples of one distorted greeble per family compared to the undistorted variant. Distortions were created using Photoshop CS6®, using 2D depictions of the greebles.

Finally, six averaged greebles (one per family, one across all greebles) were created by morphing a pair of the normal greebles of each family, and morphing the result with the morph between a different pair of greebles of the same family. The morphing result was then morphed with the final member of the family with an 80:20 weighting to create the family average. Finally, pairs of family averages were morphed in the same manner to create a total averaged greeble. 70–100 morphing landmarks were used for each morphing procedure. Landmarks were positioned around greebles' main bodies, heads, and body parts, as well as along lines indicating the shapes of certain features (e.g., lower body portions which were present in some greebles). Example morphing landmarks of different family averages landmarks are shown in *Figure A3*.

A summary of the greeble families and differences between greeble conditions are depicted in *Figure 4.1*. Morphing was conducted via Fantamorph Deluxe®, and morphing noise was eliminated with Photoshop CS6®. All greeble stimuli used can be found in the supplementary files.

Figure 4.1

Two individual greebles per family, a distorted greeble of the first family, and the total average. Unedited stimulus images courtesy of Michael J. Tarr, Carnegie Mellon University, http://www.tarrlab.org/.



Procedure

Expertise training. Expertise training was based on a five-session setup successfully used in previous studies (Bukach et al., 2012; Gauthier et al., 1999). Only participants within the *training group* completed the expertise training. The different tasks are described below while the procedure is summarized in *Table A1*. The training regime took place over five days and took 60 minutes each day on average. The tasks were used to familiarise the participants with the greebles as follows: *Family examples* (2 greebles per family are presented together with the family labels in *Figure 4.1*); *Family viewing* (participants view an individual greeble with the respective family label); *Family naming* (participants view an individual greeble and must press the first letter of the greeble's family name), *Individual viewing* (participants view an individual greeble with the respective individual label), *individual naming* (participants view an individual greeble and must press the first letter of the greeble's individual name), *Individual naming with feedback* (like *Individual naming* but participants see the correct label after an incorrect response), *Verification* (participants view an individual greeble followed by either a family or individual label that is correct 50% of the time. Participants press *y* when the label is correct, *n* when it is incorrect, and *space* if the greeble or label have not yet been shown

before), and *Final verification* (same as verification; the accuracy data was used to check whether participants acquired expertise). Each task had a fixed number of trials which are summarized in *Table A1*.

Rating task. The rating task was completed by participants of both conditions. *Training condition* participants completed the ratings task after the final expertise session whereas the *control condition* participants had not previously seen any greebles when completing the rating task. Participants rated normal, averaged, and distorted greebles on seven scales ranging from 0 to 100: *eerie*, *creepy*, *strange*, *weird*, *pleasant*, *attractive*, and *appealing*. Based on previous research on the categorization of measures of the uncanny valley (Diel et al., 2022), *eerie* and *creepy* are combined to an *uncanniness* index, *strange* and *weird* to an abnormality index, and *pleasant*, *attractive*, and *appealing* to an *attractiveness* index. *Uncanniness* thus reflects a negative specific emotional reaction, *abnormality* a judgment of the stimulus' atypicality or unusualness, and *attractiveness* the stimulus' aesthetic appeal, all which are related to an uncanny valley (Diel et al., 2022). Participants rated a total of 41 greebles (25 normal greebles, 10 distorted greebles, 6 average greebles).

Analysis, ethics statement, and data availability

Data preparation, data cleaning, and statistical analysis were conducted via JASP and R. Main analysis and post-hoc tests were done via Bayesian mixed-effects analyses of variance (ANOVA) in JASP. After sampling a first set of participants, BF >3 was used as a threshold for the Bayesian stopping rule. Bayesian mixed-level ANOVAs were all conducted with the same default options on JASP (Prior r scale fixed effects = 0.5, prior r scale random effects = 1. Further specifications to reproduce the results are available on OSF: https://osf.io/zsnkr/). Fixed and random effects were defined for each analysis (see openly available source code for JASP for how formulas are defined; Love et al., 2019). Accompanying non-Bayesian

ANOVAs and post-hoc tests were done with R, and linear mixed models were used for post-hoc tests. Linear mixed models produce large degrees of freedom seen in the results section (Kuznetsova et al., 2017; Luke et al., 2017). R packages *ez* (function *ezANOVA()*)and *nlme* (function *lme()*)were used for ANOVAs and linear mixed models, respectively.

The study was approved by the Cardiff University School of Psychology Research Ethics Committee in March 2021 (reference number: EC.21.02.09.6291R). The data and R code for the analysis is available at: https://osf.io/zsnkr/.

Evaluation of the results is based on the size of the BF as recommended (Dienes, 2021; Van Doom et al., 2020): $BF_{10} < 1$ is interpreted as no evidence for the alternative hypothesis, $1 < BF_{10} < 3$ as weak evidence, $3 < BF_{10} < 10$ as moderate, and $BF_{10} > 10$ as strong evidence. BF have been prioritized over p-values in evaluating one hypothesis over another, as there are limitations in interpreting significance of p-values (Krypotos et al., 2017).

Results

Expertise acquisition

Because an accuracy of 33% in the final verification chance would indicate a random chance response, only trained participants with an accuracy above 40% were included in the analysis. The 40% threshold was selected to exclude the potential of random chance responses. One of the 21 participants in the expertise condition had an accuracy below the threshold (35%) and was thus excluded from the analysis. The remaining participants' average accuracy in the final verification task was $M = 70.29\%$, $SD = 14.94\%$.

Greeble rating

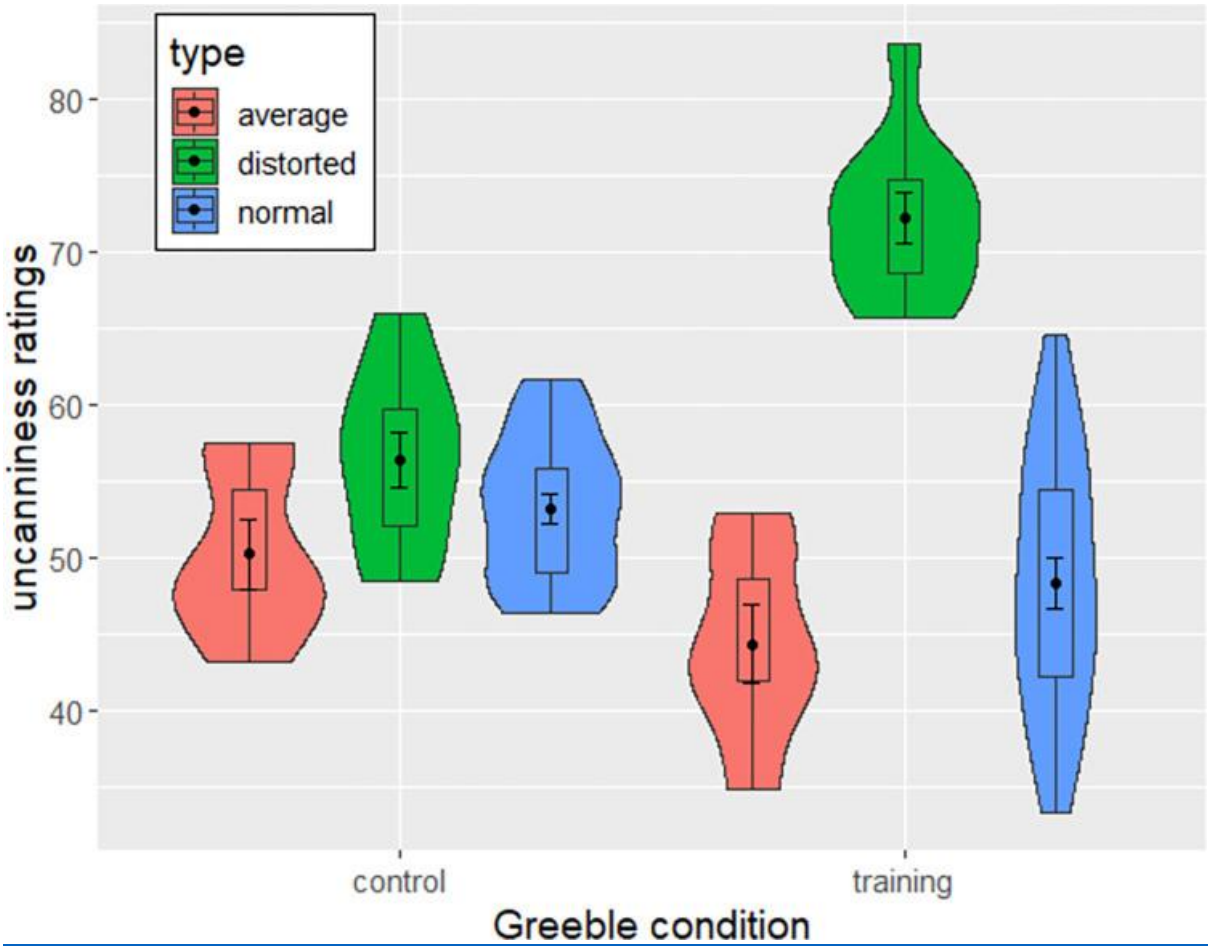
Rating scales. The items *eerie* and *creepy* were combined into an *uncanniness* index, the items *strange* and *weird* into an *abnormality* index, and the items *pleasant*, *attractive*,

and *appealing* into an *attractiveness* index. Indices were averages of the items across data. Outlier values of *uncanniness* and *attractiveness*, defined as +/- 1.5 of the interquartile range from the median, were detected and removed for all 2x5 conditions (three outlier values in total). Across all participants, *uncanniness* had a Cronbach's alpha of α = .88, *abnormality* had α = .82, and *attractiveness* α = .87, indicating good reliability for all three indices.

Uncanniness and abnormality ratings. The mean uncanniness ratings for each type of greeble for expertise and control participants are shown in *Figure 4.2*. A mixed-effects ANOVA with greeble type (average, normal, or distorted) and condition (training or control) as predictors of uncanniness ratings and participants as within-subject variables showed strong evidence for a main effect of greeble type (BF_{10} = 1206.401, $F(2, 80) = 34.528$, $p < .001$, $\eta^2 = .16$) but no main effect of condition (BF_{10} = 0.307, $F(1, 40) = 0.086$, $p = .770$, $\eta^2 < .01$). However, there was strong evidence for an interaction between type and condition (BF_{10} = 4.101e^{12}, $F(2, 80) = 19.559$, $p < .001$, $\eta^2 = .09$) on uncanniness ratings. The same pattern was observed for abnormality ratings (main effect type: BF_{10} = 508.565, $F(2,80) = 35.067$, $p < .001$, $\eta^2 = .11$; main effect condition: BF_{10} = 0.361, $F(1,40) = 0.124$, $p = .726$, $\eta^2 < .01$; interaction condition and type: BF_{10} = 7.598e^{11}, $F(2,80) = 15.471$, $p < .001$, $\eta^2 = .05$).

Figure 4.2

Violin and boxplots depicting uncanniness ratings across greeble types and conditions.



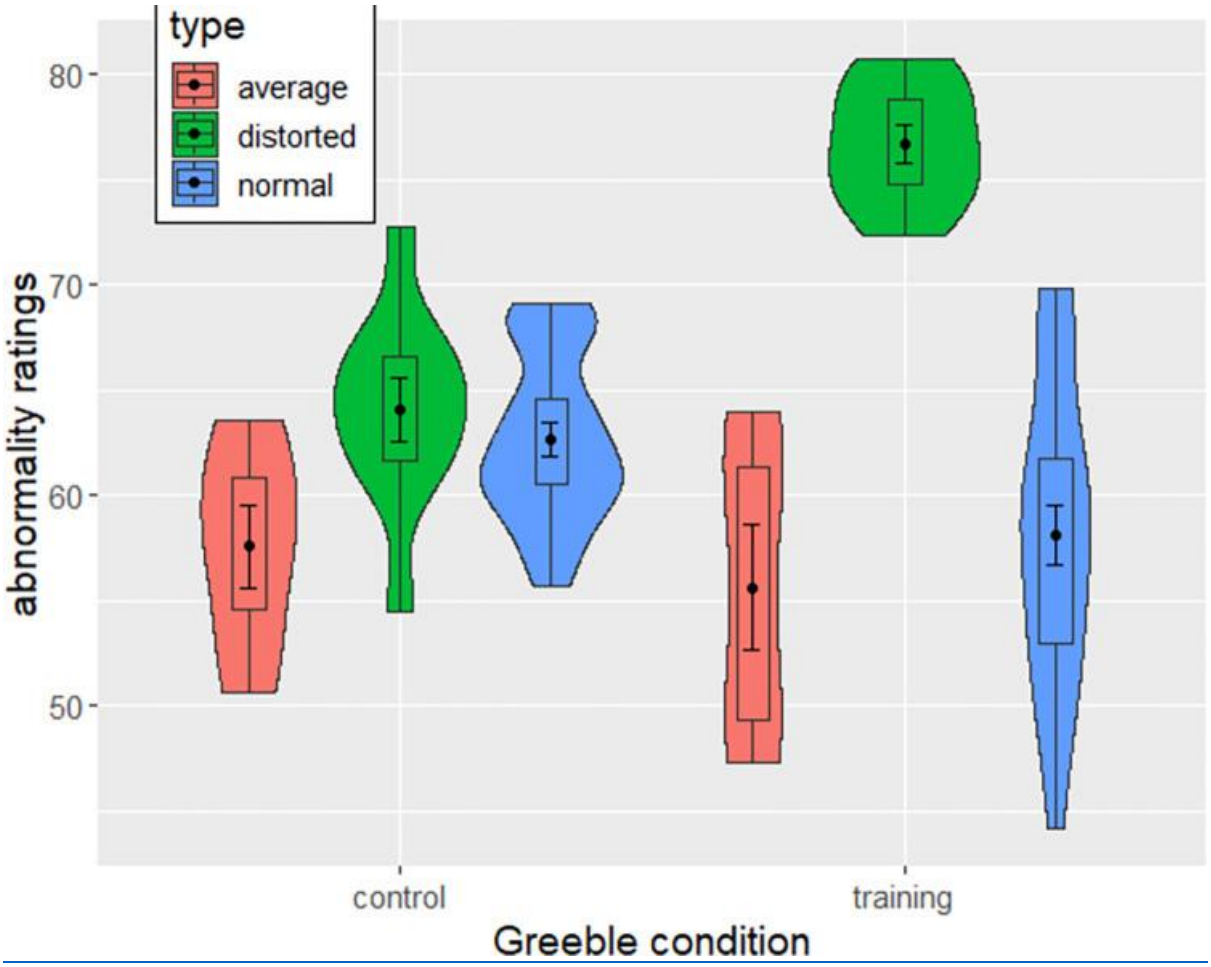
P-adjusted post-hoc Tukey tests were conducted to further investigate the interaction between greeble type and condition. Linear mixed models were used for non-Bayesian post-hoc tests, with greeble types and condition as fixed effects and participants as random effects. Comparisons showed no evidence in favour of distorted greebles being more uncanny than normal greebles without the training (BF_{10} = 0.279, t (1501) = 1.711, p_{adj} = .488, d = .14), there was strong evidence that distorted greebles were more uncanny than normal greebles after expertise training (BF_{10} = 1.894e^{20}, t(1501) = 13.399, p_{adj} < .001, d = 1.2). Most interestingly, there was strong evidence that distorted greebles were more uncanny after training than distorted greebles in the control condition (BF_{10} = 20959.55, t(1501) = 2.909, p_{adj} = .027, d = 0.66).

Finally, there was no evidence that averaged greebles were less uncanny than normal greebles in the control condition (BF_{10} = 0.168, t(1501) =, p_{adj} = .311) nor in the training condition (BF_{10} = 0.366, t(1501) = -1.431, p_{adj} = .153).

Abnormality ratings follow a similar pattern with strong evidence in favor of the model: Distorted greebles were not more abnormal than normal greebles without training (BF_{10} = 0.12, t(1501) = 0.719, p_{adj} = .98), but they were more abnormal after the training with strong evidence (BF_{10} = 5.363e^{12}, t(1501) = 11.519, p_{adj} < .001, d = 1.03). In addition, there was strong evidence that distorted greebles were more abnormal after training than without training (BF_{10} = 78.266, t(1501) = 1.875, p_{adj} = .372). Ratings are depicted in *Figure 4.3*.

Figure 4.3

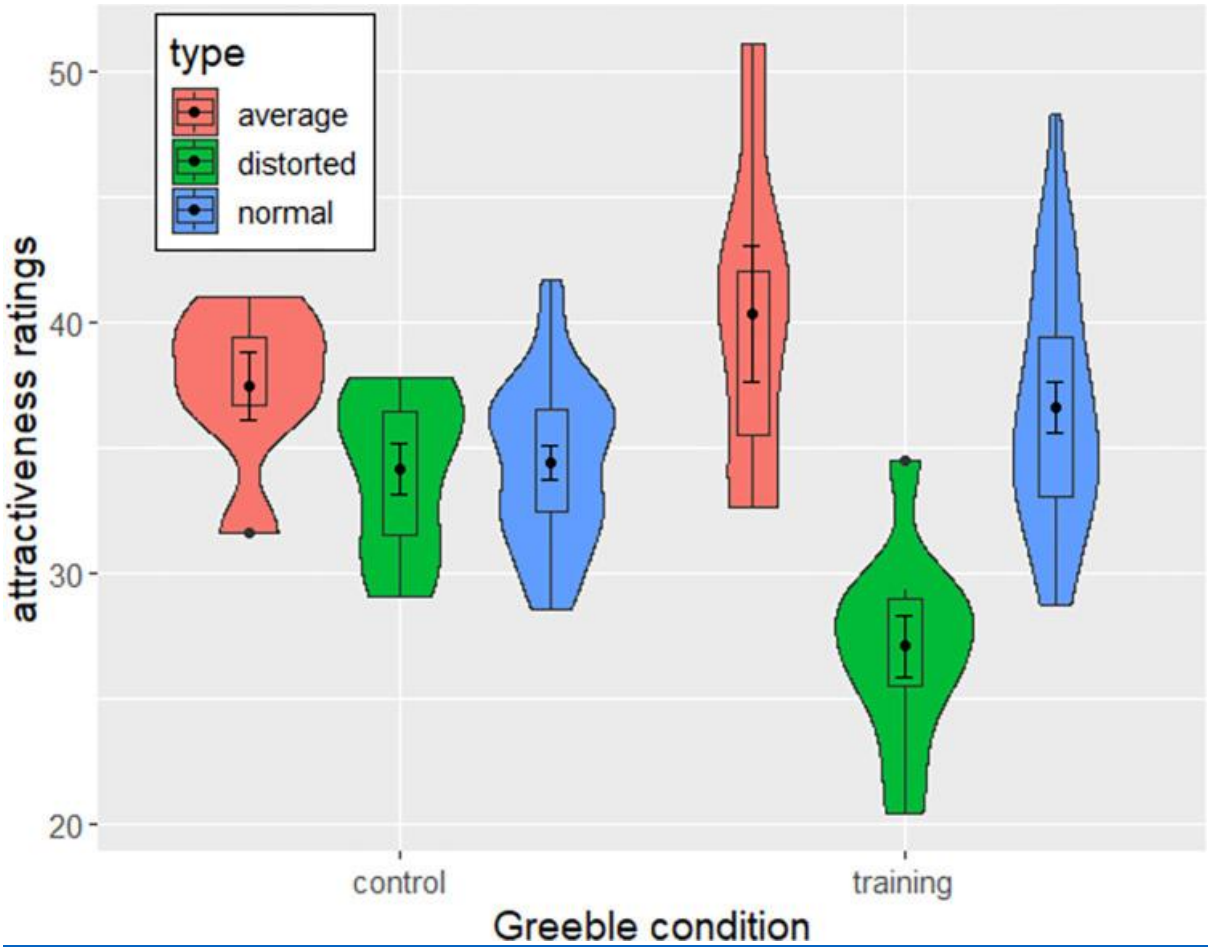
Violin and boxplots depicting abnormality ratings across greeble types and conditions.



Attractiveness ratings. *Figure 4.4* depicts the mean *attractiveness* ratings for the three greeble types of the expertise and control conditions. Similarly to *uncanny* ratings, the ANOVA with greeble type and condition as predictors of *attractiveness* ratings showed moderate evidence for a main effect of greeble type (BF_{10} = 4.46, $F(2, 80) = 25.184$, $p < .001$, $\eta^2 = .06$) but not for a main effect of condition (BF_{10} = 0.299, $F(1, 40) = 0.03$, $p = .861$, $\eta^2 = .00$), and strong evidence for an interaction effect between greeble type and condition (BF_{10} = 6705.494, $F(2, 80) = 8.707$, $p < .001$, $\eta^2 = .02$) on *attractiveness* ratings.

Figure 4.4

Violin and boxplots depicting attractiveness ratings across greeble types.



Post-hoc Tukey tests furthermore showed no evidence that average greebles were more attractive than normal greebles in either the control condition (BF_{10} = 0.334, $t(1463)$ =

2.055, p_{adj} = .04, d = 0.225), or after expertise training (BF_{10} = 0.398, t(1463) = 2.164, p_{adj} = .043, d = 0.15). However, the p-values indicated significance in both tests. Furthermore, there was no evidence that distorted greebles were less attractive than normal greebles in the control condition (BF_{10} = 0.096, t(1463) = 0.732, p = .46, d = .16), but strong evidence in the training condition (BF_{10} = 52297.5, t(1463) = -7.516, p < .001, d = .68).

Verification task accuracy as a predictor. Given the prediction that expertise modulates the effect of deviation on uncanniness, a higher degree of expertise would likely predict stronger uncanniness (and abnormality) and lower appeal ratings of distorted greebles. Thus, post-hoc regression model analyses were conducted with verification task accuracy as a predictor for uncanniness, abnormality, and appeal ratings respectively, in the training condition. Results indicate no evidence for verification accuracy as a predictor for uncanniness (BF_{10} = 0.215, t(633) = -1.566, p = .118, R^2 = 0.003), moderate evidence for predicting abnormality (BF_{10} = 3.804, t(633) = 2.932, p = .004, R^2 = .011), and weak evidence for predicting appeal (BF_{10} = 1.097, t(633) = 2.334, p = .02, R^2 = .008). Thus, verification accuracy, a proxy for the degree of acquired expertise, slightly predicted aesthetic ratings of greebles.

Changes of ratings across trials. While the main analysis provides insight into the effect of expertise on uncanniness, abnormality, and attractiveness ratings of deviating and prototypical variants, it is not clear whether expertise specifically, or a different process like general familiarization or exposure, is necessary to facilitate the difference between greeble types. In addition, according to p-values (but not Bayesian statistics), prototypical greebles were more attractive than normal greebles without expertise training. This difference may be the result of a prototypicality preference after early exposure (and thus an increase of attractiveness of averaged greebles across trials), or because of characteristics intrinsic to the stimuli (e.g., morphing artefacts). To investigate whether early expose affects the ratings of greebles (e.g., an early increase of average greebles by the end of the rating task), post-hoc

analyses on the greeble type ratings were conducted depending on the greebles' trial occurrence, for both the control and training condition. Thus, the post-hoc analysis aims to investigate an effect of early familiarity on changes in ratings of greeble types by testing for an interaction between greeble types and trial number.

Linear mixed models were used to investigate the effect of the trial on which the stimulus occurred on ratings, across greeble types and conditions, with trial, type, and condition as fixed effects and participants as random effects. However, no interaction with trials was significant for uncanniness (trial and type: $BF_{10} = 2.910e^{-4}$, $t(1414) = -1.46$, $p = .887$; trial and condition: $BF_{10} = 6.786e^{-6}$, $t(1414) = -1.36$, $p = .144$; trial, type, and condition: $BF_{10} = 1.244e^{-9}$, $t(1414) = 0.35$, $p = .726$), abnormality (trial and type: $BF_{10} = 5.262e^{-6}$, $t(1414) = -0.309$, $p = .757$; trial and condition: $BF_{10} = 8.338e^{-6}$, $t(1414) = -0.038$, $p = .97$; trial, type, and condition: $BF_{10} = 0.003$, $t(1414) = -0.416$, $p = .678$), or attractiveness (trial and type: $BF_{10} = 1.047e^{-6}$, $t(1414) = -0.143$, $p = .887$; trial and condition: $BF_{10} = 5.585e^{-6}$, $t(1414) = -0.196$, $p = .845$; trial, type, and condition: $BF_{10} = 5.022e^{-9}$, $t(1414) = 0.639$, $p = .523$).

Discussion

Greebles are a set of novel objects used here to manipulate expertise acquisition (Gauthier & Nelson, 2001). Thus, they are a useful tool to assess the effect of expertise on the aesthetic evaluation of structurally deviating objects. Here it was shown that greeble experts and greeble non-experts evaluate the uncanniness and attractiveness of deviating greebles differently.

The deviation-from-familiarity effect

In line with the hypothesis, the results show that distorted greebles were significantly more uncanny than normal greebles after expertise training, but not in the control condition. Furthermore, post-training distorted greebles were significantly more uncanny than distorted

greebles without expertise training. Thus, expertise increased the uncanniness of deviating exemplars. This *deviation-from-familiarity effect* (increased uncanniness of stimuli that deviate from a highly familiar norm) provides insight into negative evaluation of distorted exemplars, and into the uncanny valley specifically: Uncanniness of humanlike and similar entities may not be explained by mechanisms specific to certain categories of stimuli like humans and animals, such as disease avoidance, mind perception, or dehumanization, but could also occur *specifically* due to a deviation from categories of stimuli that humans have familiarized over the course of (a part of) their life, like faces and facial expressions, bodies, voices, and biological motion. The results complement previous research showing that changes in uncanniness ratings are more sensitive to distortions in familiar categories like human compared to animal faces (Diel & MacDorman, 2021; MacDorman & Chattopadhyay, 2016), realistic compared to unrealistic faces (Green et al., 2008; MacDorman et al., 2009; Mäkäräinen et al., 2014), one's own face compared to a stranger's face (Weisman & Pena, 2021), and familiar compared to unfamiliar and upright compared to inverted faces (Chapter 2). In general, the deviation from familiarity effect predicts a generality of the uncanny valley phenomenon (or uncanniness) beyond previous suggestions of human or animal specificity (Mori, 2012; Ho & MacDorman, 2010; Schwind et al., 2018), and could extend to inanimate yet familiar categories like written text or physical places, which could be explored in future research.

Uncanniness and animacy

Some theories on the uncanny valley explain uncanniness through changes in animacy perception (Appel et al., 2020; Gray & Wegner, 2012; Looser & Wheatley, 2010; Wang et al.,2020; Wang et al., 2015): stimuli may be uncanny because they straddle boundaries of animacy perception, or because they are "dehumanized" through a subtraction of animacy perception. Past research associated greeble expertise training with animacy (Cheung &

Gauthier, 2014), and it has been argued that greebles already look animate (Kanwisher, 2000). Animacy was not measured in this study. Distorted post-expertise greebles as used here may have been uncanny because they approximated post-expertise normal greebles (which could be perceived as animate), yet deviated from them and thus were either perceived as ambiguously animate, or were subtracted animacy. In this sense, uncanniness may not be the result of deviation from familiarity per se, rather than from anomalies in animacy attribution (which, in turn, would result from deviations from familiar patterns). Alternatively, greebles may be perceived as animate regardless of expertise level, and animacy combined with deviation may elicit uncanniness.

Attractiveness and deviation

Similar to the pattern observed for *uncanniness* ratings, distorted greebles were less attractive than normal greebles but only on the training condition. Thus, the *deviation-from-familiarity effect* is not specific to uncanniness but can be applied to negative aesthetic evaluations in general, beyond experiences of uncanniness (e.g., ugliness). A negative correlation between uncanniness and attractiveness or likability has been demonstrated in previous research (Destephe et al., 2015; Ho & MacDorman, 2010). Experience of negative affect may decrease perceived attractiveness or likability of an artificial entity, but so can structural deviation: Ugly and Botox (and thus highly distinctive) faces are more creepy than normal faces (Olivera-La Rosa et al., 2021). Thus, evaluations of attractiveness and uncanniness may have similar underlying processes of deviation detection.

One approach to decorrelate the negative association between ratings of attractiveness and uncanniness is by using stimuli that are both attractive and uncanny: For example, sex dolls or sex robots are anecdotally uncanny or creepy and simultaneously sold because of their sexual appeal. Exaggerated sexual features and averaged faces that coincide with human deviation, like a lack of facial details and rigid poses and social responses, may thus elicit

both sexually attracting and uncanny reactions. In that case, effects of prototypicality and deviation may be combined across multiple dimensions and modalities, where prototypicality in one dimension and deviation in another can elicit a mix of attractiveness and uncanniness, implying that those constructs can be distinguished.

Why are deviating exemplars uncanny?

Repeated view of normal greebles during the training could have increased their positive evaluation while not affecting the rating of unfamiliar, distorted greebles (Zajonc, 1968). Mere exposure, specifically a lack of exposure for uncanny stimuli, has been proposed as an explanation of the uncanny valley (Burleigh & Schoenherr,2015). Increased exposure of multiple stimuli that are grouped based on structural similarity could lead to *blending* and thus a preference to prototypicality (Carr et al., 2017). However, training also *increased* the uncanniness ratings of distorted greebles despite their similarity to normal greebles. Thus, the observed *deviation-from-familiarity effect* cannot be explained by a mere exposure effect of non-deviating stimuli.

Expertise strengthens the ability to detect differences between individual exemplars. Thus, hypothetically, expertise could have amplified small pre-existing rating biases between greebles, rather than facilitating a normal categorical variation (e.g., distorted greebles could have been slightly, but not significantly, more uncanny than normal greebles even without expertise due to factors like morphing noise. Expertise would then enhance the uncanniness difference between greeble types). However, a small attractiveness bias towards averaged greebles is present both without and with expertise training with comparable effect sizes (d = 0.225 and d = 0.15). Furthermore, distorted greebles did not differ from normal greebles in the control condition on any rating, but they did in the training condition. Thus, the data do not indicate any biases that have been amplified by expertise training.

Novelty avoidance proposes that stimuli are uncanny because they are novel (Kawabe et al., 2017). While distorted greebles were presented to the training condition participants for the first time and thus would be relatively more novel than the normal greebles, post-training distorted greebles are still *less* novel than distorted greebles in the control condition. Despite being more familiar, post-training distorted greebles were less uncanny than in the control condition. In addition, averaged greebles were also presented for the first time, but were not more uncanny than normal greebles in the training condition. Thus, novel greebles were not automatically uncanny, and deviation from the familiar variance can explain the results better than novelty avoidance. In general, it seems that the proximity of a novel stimulus to a familiar pattern is integral to the observed effect: deviating greebles were only uncanny after a normal variation of greebles has been experienced. This deviation from familiarity effect could be explained by cognitive disfluency, as the distance between a familiar pattern and the deviating variant could elicit disfluent processing. Alternatively, the discrepancy between a learnt pattern and a deviating exemplar could elicit a prediction error, which has been proposed to underlie the uncanny valley in previous research (Saygin et al., 2012): The recognition of the general shape of a greeble could create a mental predictive model of the greeble after training; however, the deviating features would then violate the more specific predictions of the greeble's appearance (i.e., the position of the body parts), and elicit discomfort.

Further questions

This study's results raise multiple questions for future research. First, distorted greebles in this study always consisted of the same pattern of distorted configuration. However, distortion can vary greatly both quantitatively (degree of distortion across one dimension) and qualitatively (distortions across different dimensions). Future research can, for example, investigate how the degree of distortion influences uncanniness ratings. Similarly, it would be

interesting to see how the amount of experience influences the sensitivity to distortions: At what point during the expertise training does the *deviation-from-familiarity effect* (and the preference for prototypicality) occur, does it increase with prolonged experience, and does it get more sensitive for subtler distortions? A post-hoc analysis found that verification accuracy (a proxy for acquired expertise) did not significantly predict the greeble uncanniness ratings, but did predict their abnormality and attractiveness ratings, indicating that the acquired level of expertise does influence aesthetic ratings. A potential mechanism is that a higher level of expertise makes deviations from prototypical appearances more apparent (Chapter 2) leading to negatively experienced cognitive disfluency (Reber et al., 2004). Future research can investigate the effect of the acquired level of expertise on aesthetic ratings of deviating variants more thoroughly, for example by manipulating the duration and intensity of expertise training.

Second, investigating neural correlates of the *deviation-from-familiarity effect* and its development is of interest to better understanding the effect. According to the cognitive fluency theory, distorted greebles should elicit stronger activation patterns than normal greebles in greeble-selective brain areas after expertise training. Furthermore, it would be interesting to investigate whether post-training distorted greebles elicit prediction errors like those observed in prior research on the uncanny valley (Saygin et al., 2012).

Third, this study did only investigate the affective and aesthetical judgment of the greebles, neglecting possible cognitive components. Future research can look at cognitive mechanisms underlying the processing of greeble stimuli like categorization difficulty (Cheetham et al., 2014; Weis & Wiese, 2017; Yamada et al., 2013), distortion sensitivity (Chapter 2), configural processing (Diel & MacDorman, 2021), and whether expertise itself is necessary for the *deviation-from-familiarity effect* or if prolonged experience alone is sufficient.

Fourth, as the goal of this study was to investigate the effect of deviation and expertise on uncanniness *in principle*, only one distortion type and degree was used. Further research can investigate the interaction between the degree of deviation or distortion, and expertise, as well as the type of distortion. Previous research has shown that in faces, increasing distortions are perceived as increasingly more uncanny (Chapter 2); a similar relationship may be predicted for greebles.

Fifth, this study did not measure the perception of animacy of greebles. As animacy has been both associated with the uncanny valley (Wang et al., 2020) and greebles (Cheung & Gauthier, 2014), the observed effect of deviation from familiarity on uncanniness could potentially be mediated by animacy. Future research can investigate the role of animacy by measuring animacy perception or by using more object-like (compared to potentially animal-like) stimulus sets.

Finally, while it is suggested here that the observed difference between distorted and normal greebles would correspond to the portion in Mori's (2012) graph after the valley drop into uncanniness towards full human likeness (analogous to how distorted yet realistic faces would fall into the valley and then increase towards full human likeness with decreasing distortion). However, this study did not replicate a "proper" uncanny valley curve. The observed effects are similar to those found in Chapter 2 using face stimuli, which could be plotted along an uncanny valley function, finding that with increasing face realism, the sensitivity to distortion increased. Future research could attempt to replicate an uncanny valley of greeble by using greebles of different realism levels (e.g., normal greebles and abstract drawings of greebles).

In summary, Chapters 2 to 4 provided evidence of the statistical link between specialization and distortion sensitivity. However, except for Chapter 4 which presented research with

greeble stimuli, mostly face stimuli were used. Yet a link between specialization and distortion sensitivity should be independent of stimulus category. One category for which levels of specialization have been observed is written text. Hence, Chapter 5 will extend the previous findings and apply them on written text stimuli.

Chapter 5: The uncanniness of written text is explained by configural deviation and not by processing disfluency

Methods, experiments, and large portions of the introduction and discussion in this chapter have been published in the journal *Perception* (Diel & Lewis, 2022c).

Introduction

The previous three chapters validated a moderated function of specialization on the uncanniness of distortions in faces and novel Greeble objects. Although this refined model is capable of explaining the mechanisms underlying uncanny valley, its theoretical basis predicts that the effects should not be restricted to human(-like) stimuli and instead can be applied to any object category with a measurable degree of specialization. This chapter presents research testing the refined model in written text stimuli on various levels of specialization. Furthermore, the refined theory is contrasted to ambiguity-based explanations of the uncanny valley. As both theories are domain-independent, results should be found for written text stimuli.

Uncanniness and Processing (Dis-)Fluency

Some researchers propose that the uncanniness of entities deviating from the human norm stems from the processing disfluency elicited by categorization difficulty (Yamada et al., 2013; Carr et al., 2017). Cognitive fluency theory predicts that prototypical stimuli are easily processed and thus appealing (Halberstadt & Winkielman, 2013; Oppenheimer, 2008; Winkielman et al., 2003). Ambiguous stimuli however lead to processing disfluency, which elicits negative affect (Halberstadt & Winkielman, 2014). Context mediates processing disfluency's effect: ambiguous faces are rated negatively only when the task is to categorize them on their dimension of ambiguity (e.g., androgynous faces were rated more negatively after subjects had categorized the face as either female or male; Halberstadt & Winkielman, 2014; Owen et al., 2016; Winkielman et al., 2015). Similarly, attending to the human-likeness

dimension of androids increases androids' uncanniness (Carr et al., 2017), indicating that attending to the stimulus' ambiguity increases the effect of processing disfluency, which then enhances uncanniness.

However, low processing fluency does not always decrease the aesthetics evaluation of stimuli (Jakesch et al., 2013). Furthermore, the most categorically ambiguous stimuli on a human likeness axis are not necessarily the most uncanny (MacDorman & Chattopadhyay, 2016; Mathur et al., 2020). Although the humanoid stimuli used in MacDorman and Chattopadhyay (2016) and Mathur et al. (2020) were categorized on whether they were human or not, they may have been ambiguous on other dimensions, eliciting ambiguity-driven uncanniness. However, as previous research indicates that ambiguity should only play a role when the relevant ambiguous dimension was previously attended to (Carr et al., 2017), other ambiguous dimensions should not play a role if participants were asked to categorize the stimuli on whether they are human or not. Nevertheless, further research points towards an association between categorization difficulty and eeriness (Ferrey et al., 2015; Kawabe et al., 2017). In sum, research findings are inconsistent, and the relation between ambiguity-based disfluency and uncanniness remains unclear.

Uncanniness and Deviation From Specialized Categories

Chapters 2 to 4 provided evidence that sensitivity to deviations in specialized categories can cause uncanniness, especially faces. As configural processing of faces is thought to be mediated by experience differentiating faces based on configural patterns (Diamond & Carey, 1986), a specialization on a stimulus category would sensitize the processing system to detect even slight deviations from the typical configuration (see also Gauthier & Nelson, 2001; Tanaka & Gauthier, 1997).

Uncanniness would then be elicited by the relative atypicality of a stimulus depending on its distance to the acceptable variation of exemplars within a category. Uncanniness would further increase with the degree of familiarity to the category's typical variation. It need not depend on processing disfluency caused by the stimulus' (categorical) ambiguity.

Thus, uncanniness arising from deviations in familiar or specialized categories would be expected in various categories and most easily found in domains of higher familiarity and configural processing. Written text is one such domain, which will be explored next.

Deviation From Specialized Categories and Perceptual Disfluency

While processing fluency has been previously linked with the uncanny valley as an ambiguity-driven explanation (Carr et al., 2017), processing fluency has also been associated with a statistical occurrence (hence, typicality) of a stimulus, potentially linked to a decreased processing cost (Ryali et al., 2020). Processing disfluency would then relate less to categorical ambiguity rather than with the statistical atypicality of a stimulus based on its deviation from the prototypical appearance, for example in faces (Dotsch et al., 2016). Furthermore, it has been recently proposed that stimulus judgement is affected by the specific type of (dis-)fluency (*fluency-specificity hypothesis;* Vogel et al., 2020; see also Vogel et al., 2018): For example, disfluency of written text on a conceptual or semantic level influences truth estimation more than aesthetic appeal did, while the opposite pattern was observed for written text disfluent on a perceptual level.

Thus, ambiguity-based conceptual disfluency elicited by a stimulus may not have the same effect as perceptual disfluency caused by the stimulus' deviation from the learnt typical appearance, with the latter more likely to influence aesthetic appeal of a stimulus. In relation to the uncanny valley, uncanniness could thus be caused by disfluency created through increased processing need for deviating stimuli, regardless of whether these stimuli are

categorically ambiguous. Thus, perceptual, not ambiguity-driven, disfluency, may underlie uncanniness.

The effect of perceptual disfluency depends on the expectations towards typical appearance, which may be driven by experience (Wänke & Hansen, 2015). Given that people are more aware of deviations or changes in more familiar or specialized stimuli (Chapters 2 to 4), potential deviations may be more readily processed disfluently in those categories. Thus, the same type of deviation may appear more aesthetically unappealing in more, compared with less, specialized categories due to increased processing disfluency. In other words, the degree of familiarity or specialization would increase the sensitivity to deviations by increasing disfluency, and this effect would be more relevant for perceptual rather than for conceptual disfluency, given the specificity hypothesis (Chapters 2 to 4; Vogel et al., 2020).

Word Processing

Written words in a familiar language are recognized holistically (Pelli et al., 2003). Word and face recognition have been compared in previous research (Martelli et al., 2005) and have been associated with analogous, contralaterally aligned regions: the right fusiform gyrus for faces and the left fusiform gyrus for words and letter strings (Dehaene & Cohen, 2011; Dien, 2009; Hillis et al., 2005).

Given the similarities in word and face processing, multiple studies have successfully investigated configural processing of written words (Barnhart & Goldinger, 2013; Björnström et al., 2014; Gauthier et al., 2006; Wong et al., 2010) and its disruption in dyslexia (Conway et al., 2017). Recently, Wong et al. (2019) found that participants are sensitive to even slight changes in a word's configuration (e.g., slightly misaligning Latin letters or parts of a Chinese character), but only when they were familiar with the language and when words were presented upright instead of inverted, as inversion disrupts configural processing of stimuli

that are typically experienced upright. As observers are sensitive to subtle changes in configural patterns of words, they should also be sensitive to the uncanniness of configural word deviations if deviation from specialized categories were to cause uncanniness.

Positive effects of processing fluency on word and sentence judgement have been previously observed; for example, rhyming statements are perceived as more truthful (McGlone & Tofighbakhsh, 2000), and regular words are perceived as more familiar (Whittlesea & Williams, 1998). According to the processing disfluency hypothesis, disfluent words or sentences should elicit negative evaluation, specifically uncanniness.

Perceptual Word Disfluency

Low-level perceptual processing fluency of words can be decreased by impairing readability of sentences, for example, by using unclear fonts or decreasing contrast (Reber et al., 2004). Increased perceptual word fluency makes written information more trustworthy (Shah & Oppenheimer, 2007) and decreases the perceived distance between the reader and the stimulus (Alter & Oppenheimer, 2008), potentially by reducing heuristic processing (Alter et al., 2007). If perceptual disfluency alone decreases the aesthetic judgement, any manipulations of words or sentences decreasing their readability would then also decrease their positive evaluation.

Given an expertise-based configural processing of words, deviations from the typical configuration of words should increase perceptual disfluency, and more so for words written in familiar languages. This high-level perceptual disfluency would fit the prediction that uncanniness is caused by deviations in specialized categories.

Conceptual (Semantic) Word Disfluency

Conceptual (semantic) processing fluency may occur when the meaning of words or sentences is ambiguous (Laurence et al., 2018). Semantically ambiguous words increase

processing needs when the task is to categorize a word based on its meaning, for example within a semantic decision task (Hino et al., 2002; Piercey & Joordens, 2000; see also Eddington & Tokowicz, 2015). However, semantically ambiguous words may increase processing fluency because having multiple meanings may make them more accessible (Klepousniotou & Baum, 2007; Yap et al., 2011). Ambiguous sentences are read faster, but elicit slower processing when disambiguation is required (Logačev & Vasishth, 2016; Swets et al., 2008). As semantic categorization decreases the processing fluency of ambiguous words and sentences likely by activating competing meanings and thus a cognitive conflict, ambiguous words and sentences should be negatively evaluated immediately after a decision on their semantic meaning is required (Piercey & Joordens, 2000; Owen et al., 2016).

Experiment 6

In the present work, the effect of deviation and ambiguity on the uncanniness of written text is investigated and whether cognitive (dis)fluency or deviation from familiarity can better predict text uncanniness. The study is divided into three parts.

Research Question and Hypotheses

In the first part, the effect of familiarity on the uncanniness of configural and non-configural deviation of sentences is investigated and compared with the effect of sentence disfluency on sentence uncanniness. Sentence disfluency is operationalized as the participants' accuracy and response time for transcribing a presented sentence (*readability*). If cognitive disfluency specifically elicits the uncanniness of distorted words, stimulus manipulations decreasing fluency (readability) should also increase uncanniness:

1. Sentence readability negatively predicts the uncanniness ratings of English sentences (*disfluency*).

However, according to the theory based on deviations from specialized categories, configural deviation should increase the uncanniness of written sentences, and the effect of configural deviation specifically should increase with language familiarity.

1. Configural deviation of written sentences increases uncanniness most for a familiar language (*English*), less for an unfamiliar language that also uses Latin script (*Icelandic*), and not at all for a completely unfamiliar language and script (*Babylonian Cuneiform*). The effect of non-configural deviation (blur) on uncanniness is not affected by language familiarity (*configural deviation 1*).

In the second part, the effect of conceptual fluency (semantic ambiguity) and deviation on uncanniness is investigated. Semantic ambiguity is operationalized as the consistency of participant responses in a semantic decision task. According to the disfluency hypothesis, ambiguous words should be more uncanny after attention has been put on their semantic ambiguity:

2. Ambiguous words are more uncanny after a semantic decision task encompassing two of the words' meanings than after a semantic decision task with unambiguous answers (*conceptual disfluency 1*).

, the familiarity from deviation hypothesis would not predict an effect of conceptual disfluency effect, and instead an effect of configural deviation:

3. Configural deviations of words are rated more uncanny than non-deviating words, whether they are ambiguous or non-ambiguous (*configural deviation 2*).

Since words with ambiguous meanings may increase processing fluency due to their multiple representations rather than decreasing it (Klepousniotou & Baum, 2007), a third part of the study focussed on the effect of conceptual disfluency on uncanniness in ambiguous sentences rather than in words by investigating whether sentences with inconsistent interpretations

across participants in a sentence ambiguity task were perceived as more uncanny than non-ambiguous sentences:

4. Ambiguous sentences are rated more uncanny than non-ambiguous sentences (*conceptual disfluency 2*).
5. Configural deviations of sentences are rated more uncanny than non-deviating, ambiguous or non-ambiguous sentences (*configural deviation 3*).

Methods

Participants

According to a power analysis, 50 participants were needed to achieve a power of $1 - \beta = 0.8$. Because, to our knowledge, no study has previously investigated the effect of distortion on uncanniness, a small effect size of $d = 0.25$ was used for the power analysis (Cohen, 1988). All 50 participants were undergraduate students from the Cardiff University School of Psychology and were on average 20 years old ($SD_{age} = 1.62$) and about 96% were female.

Stimuli

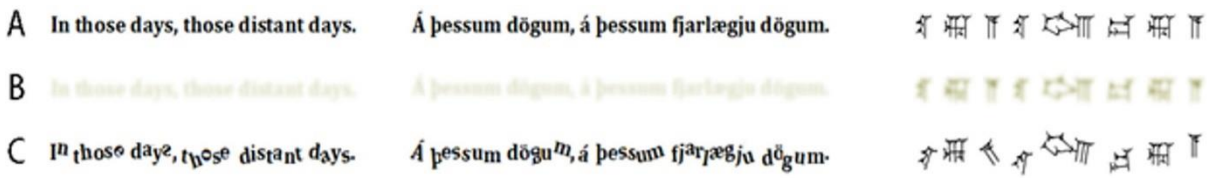
In the first part, stimuli were typical or manipulated versions of short sentences in three languages (English, Icelandic, Babylonian cuneiform). The sentences were taken from various passages of the *Epic of Gilgamesh* of the Electronic Text Corpus of Sumerian Literature (ETCSL)[1]: transliterations were transcribed into old Babylonian cuneiform using CuneifyPlus[2], and translations of the same passages were used for the English sentences. Icelandic sentences were the same passages translated by a native Icelandic speaker. A total of 15 sentences were used. For the configural distortion condition, letter and cuneiform positions and angles were changed. For the perceptual disfluency condition, sentences were blurred, and their contrasts decreased. Sentences from Babylonian literature were taken because 1) Babylonian cuneiform is guaranteed to be unfamiliar to participants, and 2)

English translations were easily available. Examples of unedited and edited sentences are shown in *Figure 5.1*, and all unedited sentences in *Table A2*.

Figure 5.1

One example sentence in English (left), Icelandic (centre) and Babylonian (right).

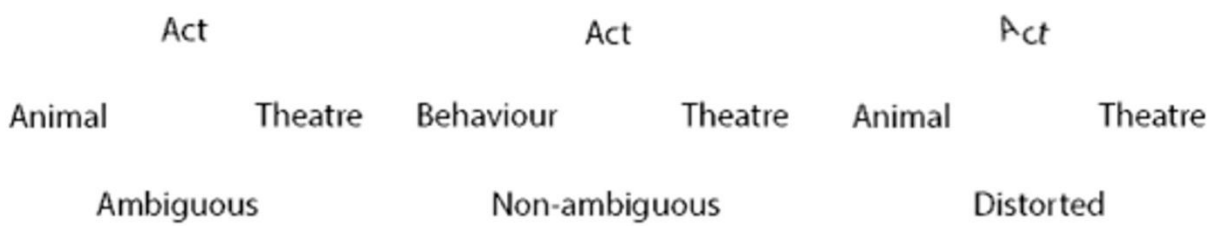
A = typical, B = blurred sentences. C = configurally distorted sentences.



For the second part, a total of 15 semantically ambiguous words were collected. Words were presented either with two other words associated with two valid meanings of the word (*ambiguity condition*), with two other words associated with only one valid meaning (*non-ambiguity condition*) or like in the non-ambiguity condition but with the word being configurally distorted identical to the distortion in the first part (deviation condition). Examples of the stimuli per condition are seen in *Figure 5.2*, and all unedited stimuli in *Table A3*.

Figure 5.2

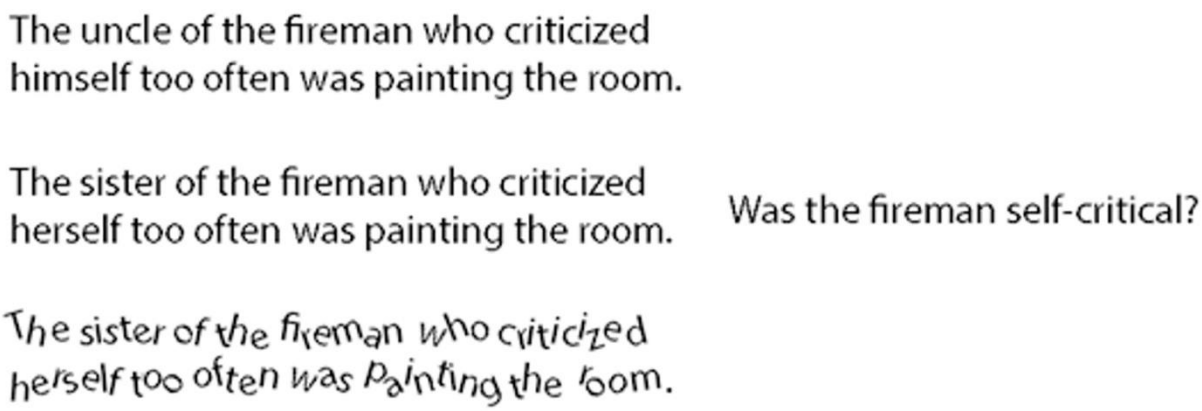
Example trials across conditions. The target word (top; here, 'Act') is presented either with two semantically associated context words (ambiguous condition), two context words of which only one is semantically related (non-ambiguous condition), or like the non-ambiguous but configurally distorted (distorted condition).



For the third part, 15 sentences have been selected which were either ambiguous (*ambiguity condition*) and had non-ambiguous counterparts (*non-ambiguity condition*). Non-ambiguous counterparts which were configurally distorted identical to the previous two parts (*deviation condition*). Sentences were derived from the selection of most ambiguous sentences (close to 50% response preference in the ambiguous condition) and non-ambiguous variants in the study by Swets et al. (2008). Example sentences for each condition are seen in *Figure 5.3*.

Figure 5.3

Example stimuli used in the final part of the study. On the left (up to down), an ambiguous sentence, a non-ambiguous sentence and a non-ambiguous-distorted sentence. On the right, the question asked on how participants interpreted the sentences.



Design and Procedure

In summary, the study was divided into three independent study tasks: A readability and rating task (parts 1a and 1b), a semantic decision and rating task (part 2) and a sentence ambiguity and rating task (part 3). The readability task followed a 3 × 1 design varying text display (normal, blur, deviation), while the rating task in task 1 followed a 3 × 3 design with both text display and language (English, Icelandic, Babylonian) as variables. Tasks 2 and 3 were again 3 × 1 designs with varying text conditions (non-ambiguous, ambiguous, deviation). The tasks will now be further elaborated.

The study was conducted online. After giving informed consent, participants followed a link to the page where they performed the experiment. Participants were randomly assigned to one of three cross-condition groups. Cross-condition groups only differed in the conditions of the base word and sentence stimuli to avoid the repeated viewing effect from the same base stimuli appearing again in a different condition; thus, each text stimulus presented was unique. Each participant viewed five stimuli per condition. All participants took part in the parts described below.

Part 1a: Readability Task

In the readability task, participants saw English versions of the sentences which were either typical (*typical condition*), configurally distorted (*deviation condition*) or blurred and decreased in contrast (*perceptual disfluency condition*) in random order. Participants were asked to type the sentence into a text box as quickly as possible and viewed five sentences per condition which were not variants of the same sentences. Participants viewed sentences per condition, and never the same sentence in different conditions.

Part 1b: Rating Task

In the Rating task, participants viewed all sentences in the *typical*, *deviation*, and *perceptual disfluency conditions* in all languages in random order and rated them on four scales used in previous research: *uncanny*, *eerie*, *creepy* and *strange* (Diel et al., 2022). Each scale ranged from 1 to 100. Scales were presented sequentially, and simultaneously with the text stimulus. Participants had unlimited time for responding.

Part 2. Semantic Decision and Rating Task

In the Semantic Decision and Rating Task, participants first viewed an ambiguous target word accompanied by two context words to the left and right. Either both context words were semantically related to the target word (*ambiguity condition*), or only one word was

semantically related (*non-ambiguity condition*), or only one word was semantically related but the target word was configurally distorted (*deviation condition*). Participants had four seconds to decide which of the context words were semantically related by pressing either the left or right key on their keyboard. Afterwards, participants had to rate the target word on a single *eerie/creepy/uncanny* scale ranging from 1 to 100. Again, participants had unlimited time to respond. Participants viewed five words per condition, and never the same word in different conditions.

Part 3: Sentence Ambiguity and Rating Task

In the Sentence Ambiguity and Rating Task, participants viewed a sentence that was ambiguous (*ambiguity condition*), non-ambiguous (*non-ambiguity condition*) or non-ambiguous but configurally distorted (*deviation condition*). Participants had unlimited time to decide whether the sentence presented was ambiguous or not, indicating their decision by pressing the left or right key. After responding, participants then rated the sentences identical to the Rating in the second part. Participants viewed five sentences per condition, and never the same sentence in different conditions.

Analysis and Ethics Statement

Analysis was conducted in R. Linear mixed models were used to control for participants, as well as linear regressions. Data cleaning was conducted by removing all outlier (1.5*IQR) uncanniness and categorization reaction time ratings for each stimulus. Numbers of outlier values removed were 20 out of 270 (task 1), 5 out of 810 (task 2), 41 out of 450 (task 3) and 31 out of 420 (task 4). The experiment was approved by the Cardiff University School of Psychology Ethics Committee in October 2021 (reference number: EC.21.09.14.6411G). The stimuli, data and analysis are available online at https://osf.io/yt9er.

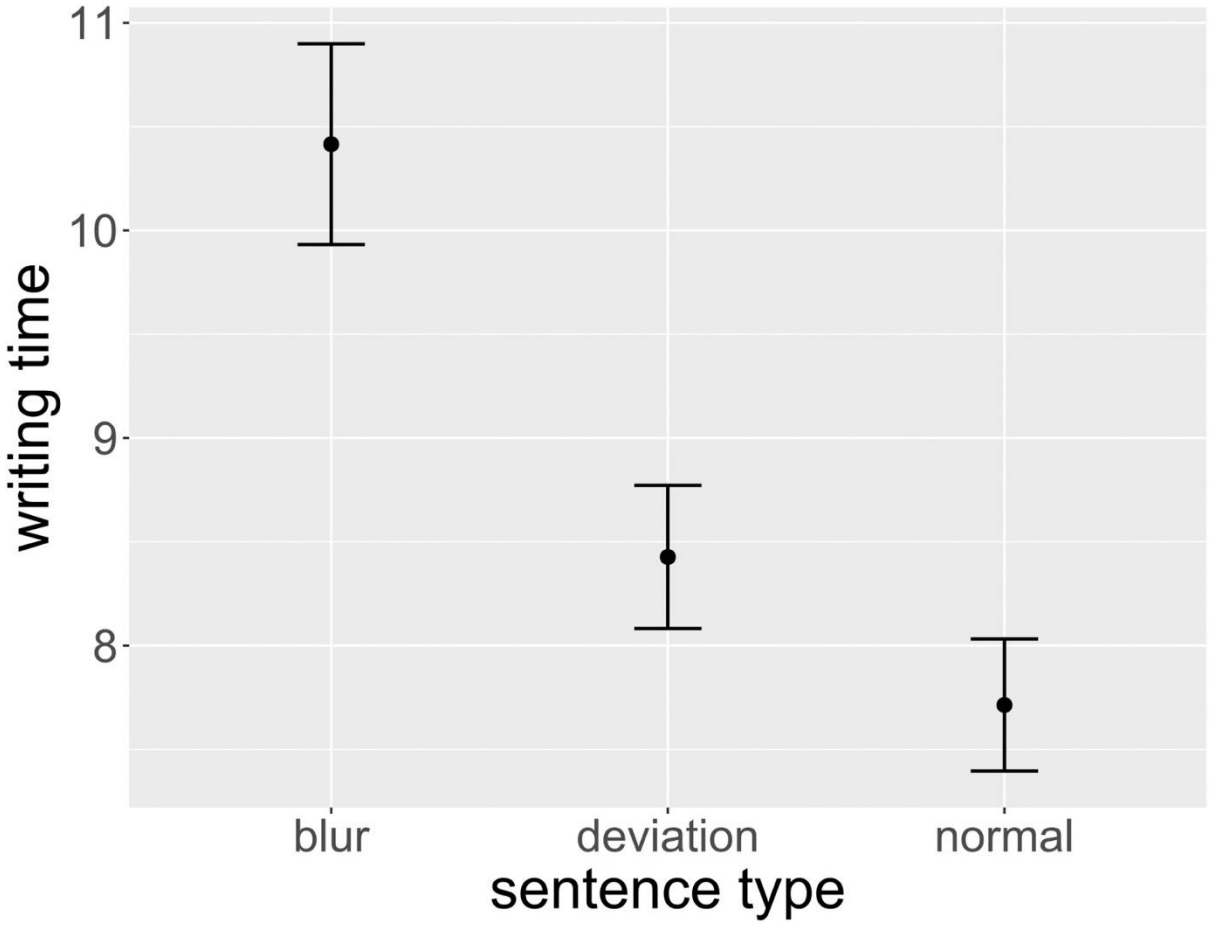
Results

Part 1. Readability, Language and Uncanniness

Sentence Readability and Uncanniness. A linear mixed model was calculated with participant and base sentence as random factors and sentence type as fixed factors. Results show a significant main effect of both blur ($t(215) = 7.36$, $p < .001$) and deviation ($t(215) = 2.15$, $p = .033$) on readability. *P*-adjusted post hoc tests revealed that while blurred sentences were significantly more difficult to rewrite than typical ($t(216) = -7.36$, $p < .001$) and deviating sentences ($t(216) = 5.25$, $p < .001$), there was no difference in readability between typical and deviating sentences ($t(216) = -2.15$, $p = .082$). The data is depicted in *Figure 5.4.*

Figure 5.4

Average time needed to replicate the sentences (in seconds) divided by sentence type. Error bars represent by-participant standard errors.



Another linear mixed model with the same random effects but readability as a fixed effect showed that reaction time significantly predicted uncanniness ($t(210) = 4.78$, $p < .001$, $R^2_{adj} = .41$). While the perceptual disfluency hypothesis is supported, it cannot explain why configurally deviating sentences are uncanny despite not being significantly more disfluent than typical sentences. Thus, perceptual disfluency cannot fully explain the results.

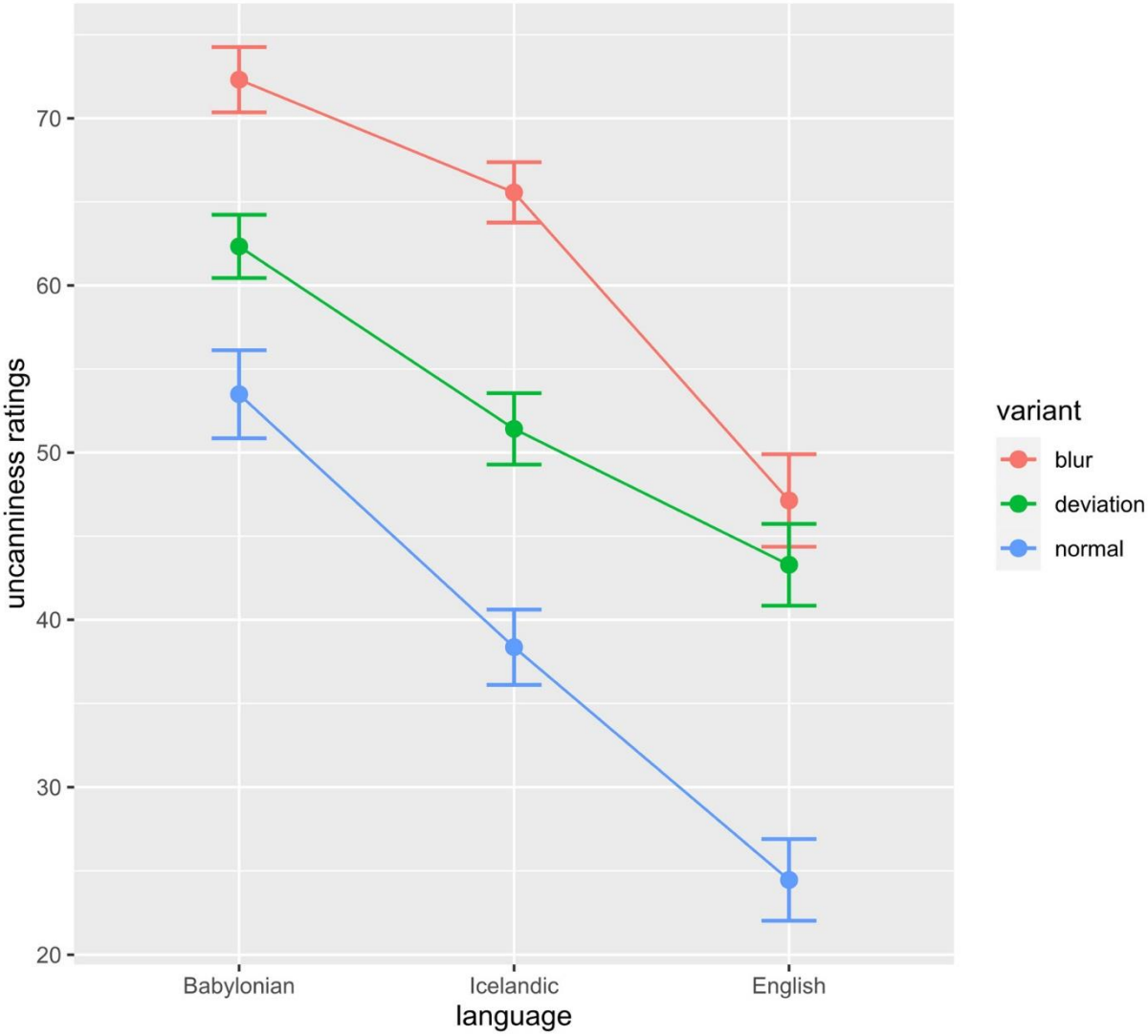
Sentence Language and Uncanniness Ratings. Sentence uncanniness ratings were tested using a linear mixed model with base sentence and participants as random effects and sentence type and language as mixed effects. Results show a main effect of language ($t(678) = -9.22$, $p < .001$), blur ($t(679) = 7.23$, $p < .001$) and deviation ($t(678) = 2.86$, $p = .004$) compared with typical. While the interaction between language and

blur was not significant, the interaction between language and deviation was ($t(678) = 2.26$, $p = .024$).

P-adjusted Tukey tests furthermore showed that for Babylonian text, blur was more uncanny than deviation ($t(676) = 3.28$, $p_{adj} < .004$, $d = 0.53$) and typical ($t676) = 5.93$, $p_{adj} < .001$, $d = 0.95$), and deviation more uncanny than typical ($t(676) = 2.69$, $p_{adj} = .033$, $d = 0.43$). Similarly, for Icelandic, blur was more uncanny than deviation ($t(674) = 4.55$, $p_{adj} < .001$, $d = 0.72$) and typical ($t(675) = 8.65$, $p_{adj} < .001$, $d = 1.36$), and deviation more uncanny than typical ($t(675) = 4.07$, $p_{adj} < .001$, $d = 0.64$). For English, blur was not significantly more uncanny than deviation ($t(674) = 1.34$, $p_{adj} = .818$, $d = 0.21$), while both blur ($t(675) = 7.22$, $p_{adj} < .001$, $d = 1.15$) and deviation ($t(674) = 5.84$, $p_{adj} < .001$, $d = 0.93$) were significantly more uncanny than typical. The data are summarized in *Figure 5.5*. Thus, the results support the deviation from familiarity hypothesis.

Figure 5.5

Average uncanniness ratings across sentence types and languages. Error bars represent standard errors.



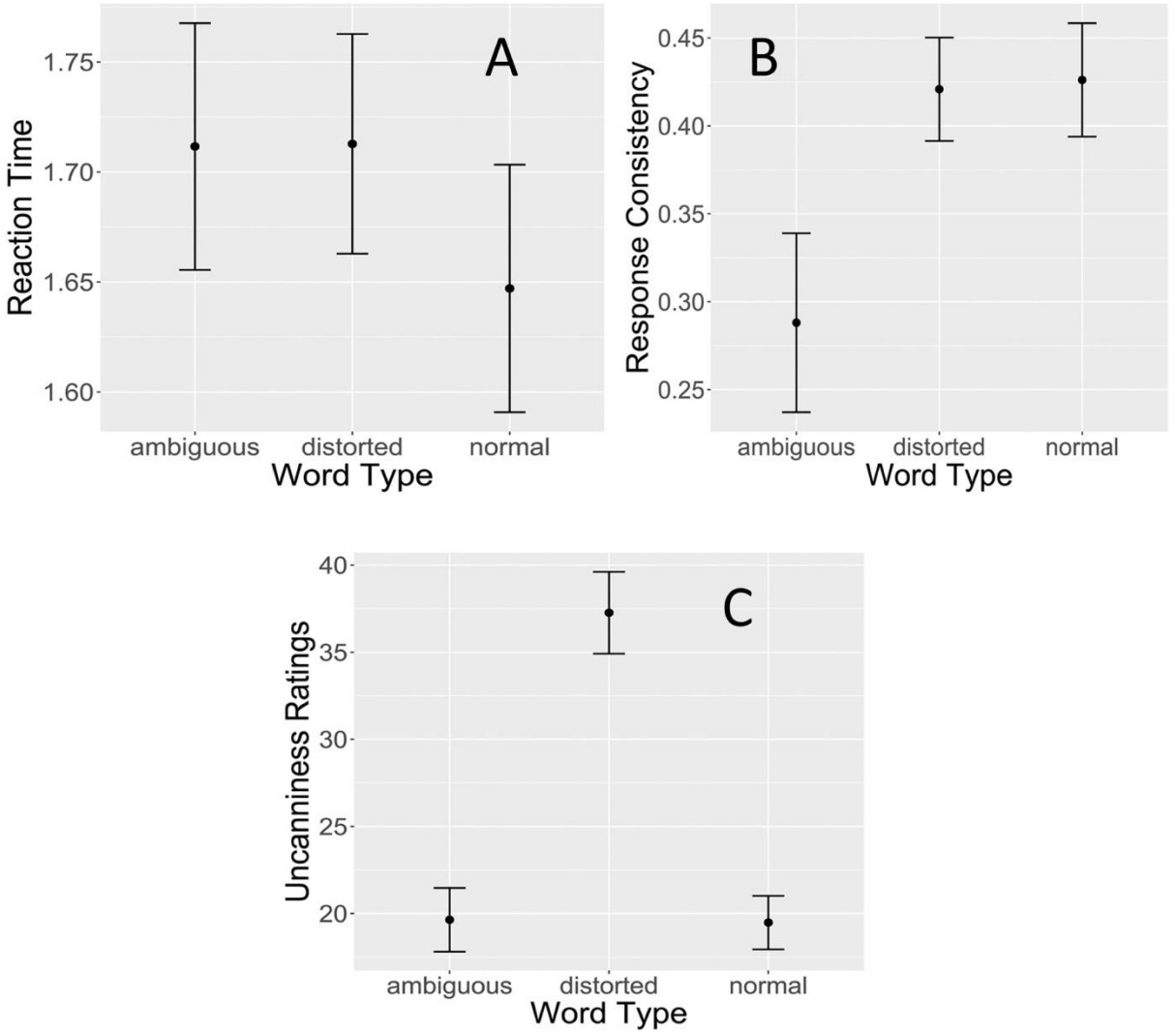
Part 2. Word Ambiguity and Uncanniness

Manipulation Check for Ambiguity. A manipulation check for ambiguity was done by comparing two indicators of categorization difficulty between word types: categorization reaction time and categorization response. Categorization responses were transformed into a categorization consistency scale, ranging from 0 (categorization at chance level) to 0.5 (consistent categorization across all participants). Linear mixed models with participants and base words as random effects and word type as fixed effects showed no effects of word ambiguity ($t(390) = 1.13$, $p_{adj} = .258$) or word distortion ($t(390) = 1.25$, $p_{adj} = .211$) on reaction time. However, word ambiguity ($t(28) = -2.32$, $p_{adj} = .028$), but not word deviation ($t(28) = -0.02$, $p_{adj} = .99$), had an effect on response consistency. Specifically, typical words

were more consistent than ambiguous words ($t(28) = 2.32$, $p_{adj} = .028$), but not deviating words ($t(28) = 0.02$, $p_{adj} = .988$), and deviating words were more consistently categorized than ambiguous words ($t(28) = -2.3$, $p_{adj} = .015$). Reaction time and categorization data are summarized in *Figure 5.6A* and *B*. Thus, the ambiguity manipulation was successful.

Figure 5.6

A: Average response reaction times across word types. B: Participants' average response consistency (0 = random, 0.5 = full consistency) across word types. C: Average uncanniness ratings across word types. Error bars represent standard errors.



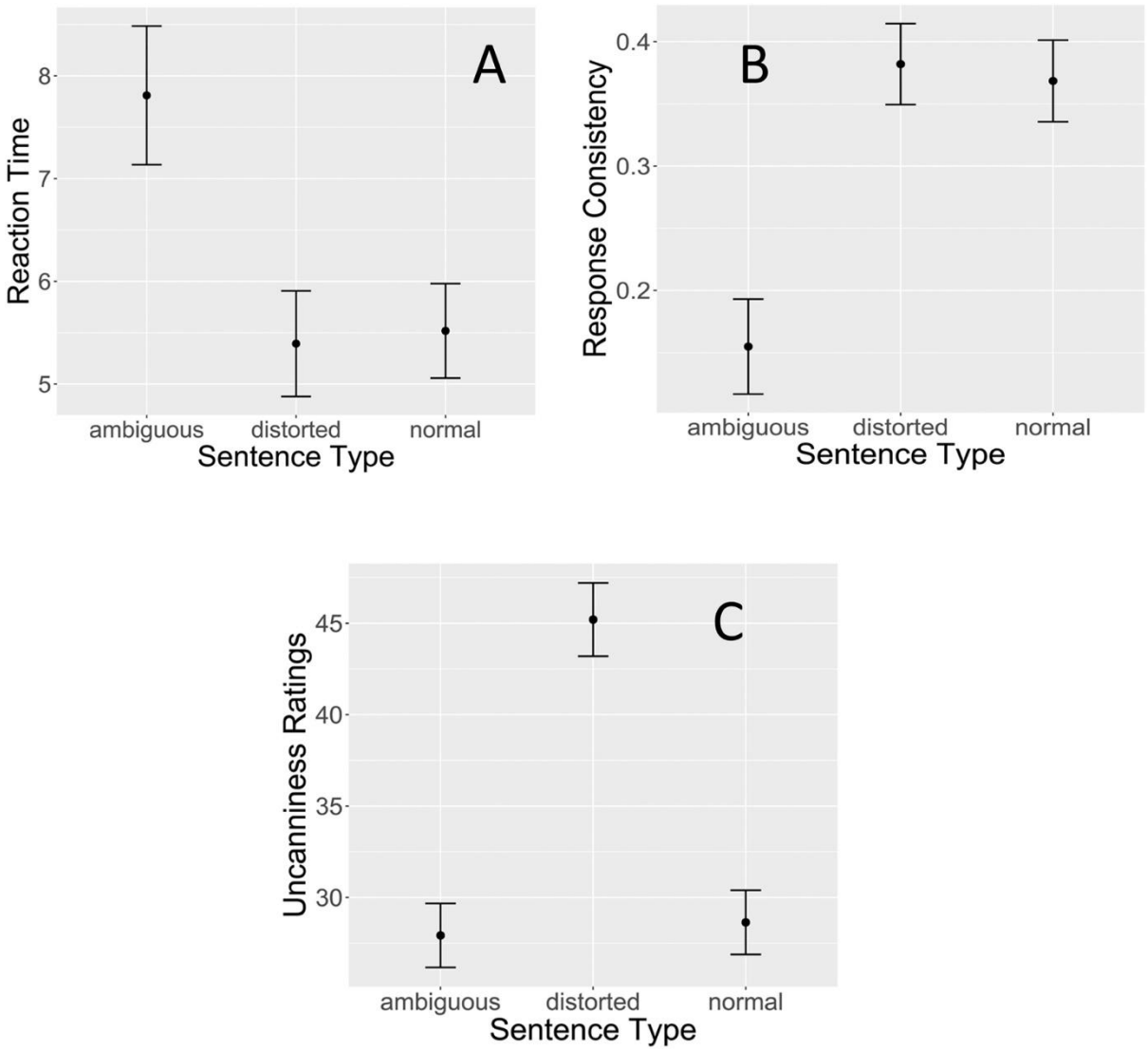
Uncanniness Ratings. Linear mixed model analysis with participants and base word as random effects and word type as fixed effect showed no effect of both word ambiguity ($t(392) = 0.02$, $p = .98$), but an effect of deviation ($t(392) = 7.86$, $p < .001$) on uncanniness. Specifically, post hoc Tukey tests showed that while typical words were not less uncanny than ambiguous words ($t(392) = -0.02$, $p = .869$), both typical ($t(392) = -7.86$, $p < .001$) and ambiguous ($t(392) = -7.84$, $p < .001$) words were less uncanny than deviating words. Data is depicted in Figure 5.6C. Thus, the configural deviation hypothesis received stronger support than the conceptual disfluency hypothesis.

Part 3. Sentence Ambiguity and Uncanniness

*Manipulation Check for Ambiguity***.** Reaction time and response consistency were used as indicators of a successful manipulation of ambiguity. Linear mixed models with participants and base sentences as random effects and sentence type as main effects showed a significant effect of sentence ambiguity on reaction time ($t(379) = 3.65$, $p < .001$), but not of sentence distortion ($t(379) = -0.11$, $p = .91$). Specifically, post hoc Tukey tests show that ambiguous sentences needed a significantly longer reaction time than typical ($t(379) = 3.65$, $p < .001$) and deviating ($t(380) = 3.8$, $p < .001$) sentences, but there was no difference between deviating and typical sentences ($t(379) = 0.11$, $p = .91$). Furthermore, response consistency analysis showed an effect of ambiguity ($t(28) = 7.19$, $p < .001$), but not deviation ($t(28) = 0.42$, $p = .676$) on consistency, and post hoc Tukey tests show that ambiguous sentences had less response consistency than typical ($t(28) = 7.19$, $p < .001$) and deviating sentences ($t(28) = 6.77$, $p < .001$), which did not differ from one another ($t(28) = -0.42$, $p = .676$). Data is summarized in *Figures 5.7A* and *B.* The ambiguity manipulation was thus successful.

Figure 5.7

A: Average response reaction times across sentence types. B: Participants' response consistency (0 = random; 0.5 = full consistency) across sentence types. C: Average uncanniness ratings across sentence types. Error bars represent standard errors



Uncanniness Ratings. A linear mixed model with participants and base sentence as random effects and sentence type as a fixed effect showed sentence deviation ($t(362) = 7.710$, $p < .001$) rather than sentence ambiguity ($t(362) = -0.14$, $p = .892$) had a significant effect on uncanniness. Post hoc Tukey tests showed that while typical sentences were not less uncanny than ambiguous sentences ($t(361) = 0.14$, $p = .911$), both typical ($t(361) = -7.71$) and ambiguous sentences ($t(362) = -7.87$, $p < .001$) were less uncanny than

deviating sentences. The data is summarized in *Figure 5.7C.* Again, the configural distortion hypothesis received support rather than the conceptual disfluency hypothesis.

Discussion

Sentence Readability and Uncanniness

The first hypothesis (*disfluency*) states that the processing fluency of sentences should increase their uncanniness. Sentence readability reaction time was used to assess participants' ability to replicate a sentence in different conditions and used as an indicator of processing fluency because impaired sentence readability increases disfluency (Reber et al., 2004). Reaction time significantly predicted uncanniness ratings. Furthermore, sentence deviation did not significantly increase reaction time, while blurred sentences were significantly harder to replicate than both typical and deviating sentences. Thus, processing disfluency seemed highest for blurred sentences while it did not show any effect for deviating sentences. However, despite having the same readability as typical sentences, deviating English sentences were significantly more uncanny than typical sentence and comparably to blurred sentences. Thus, while time needed to replicate sentences could predict uncanniness ratings, the uncanniness of deviating sentences cannot be explained by processing disfluency. Thus, the first hypothesis (*disfluency*) is partially supported.

Sentence Familiarity and Uncanniness

The second hypothesis (*configural deviation 1*) stated that the effect of deviation on uncanniness decreases as the language becomes less familiar. Specifically, deviating sentences should be most uncanny compared with typical sentences (most familiar) and least compared with Babylonian cuneiform (least familiar). Both blurred and deviating sentences were significantly more uncanny than typical sentences across languages. However, an interaction between language familiarity and deviation was observed for configurally deviating sentences, not for blurred sentences. In addition, effect sizes show that the

uncanniness difference between deviating and typical sentences increased with language familiarity from Babylonian ($d = 0.43$) to Icelandic ($d = 0.65$) to English ($d = 0.93$), which was not observed for the difference between blurred and typical sentences (Babylonian: $d = 0.95$; Icelandic: $d = 1.36$; English: $d = 1.16$). Thus, the effect of configural deviation on uncanniness decreased with decreasing language familiarity, while the effect of non-configural deviation (blur) remained constant. Thus, the second hypothesis (*configural deviation 1*) is supported.

Word and Sentence Ambiguity and Uncanniness

The third and fifth hypotheses (*conceptual disfluency 1 and 2*) stated that ambiguity increases the uncanniness of words and sentences, respectively. In contrast, the fourth and sixth hypotheses stated that configural deviation of written words and sentences increases uncanniness. Ambiguity was manipulated by adding a lexical ambiguity condition for words and a semantic ambiguity condition for sentences. A manipulation check of ambiguity (differences in reaction time and response consistency) showed partial support of successful ambiguity manipulation for words, and full support for sentences. Nevertheless, both ambiguous words and sentences were not more uncanny than typical words and sentences. Instead, non-ambiguous but configurally deviating words and sentences were more uncanny than both typical and ambiguous variants. Thus, the results indicate that configural deviation, not ambiguity, elicits uncanniness (*configural deviation 2 and 3*).

It is possible that the ambiguity manipulation in Tasks 2 and 3 could not compete with a manipulation as salient as the deviation condition, and hence was less uncanny as the deviation condition. Ambiguity was associated with aesthetic devaluation in previous research (e.g., Carr et al., 2017), but the effect may not be as strong as the effect of deviation on uncanniness. However, because the uncanniness difference between the normal and

ambiguity condition was not significant, the results of this study do not indicate any kind of effect of ambiguity on uncanniness.

Processing disfluency is a reaction relative to the expectation of an occurrence (Wänke & Hansen, 2015). Hence, the typical variation of letter structure is expected to be much narrower than the variation of the content of a sentence. Hence, the observed effect of deviation, but not ambiguity, may be because the former condition elicits greater typicality-based fluency than the latter. Nevertheless, the results suggest that ambiguity-based disfluency alone is not sufficient to explain uncanniness.

Human-Specificity of Uncanniness

Various theories predict that uncanniness results from anomalies in human-specific processing (Stein & Ohler, 2017; Wang et al., 2020). However, the face stimuli used in studies investigating human-specific processes have been variants deviating from typical facial appearance. The present work shows that anomalies deviations in specialized categories like written text can elicit uncanniness in themselves, and human-specific processes can be excluded. Given the analogous processing of written text and faces, configural atypicalities in artificial faces may thus already be uncanny because of their deviation, while also influencing later human-specific processing like dehumanization or threatening human identity. Thus, uncanniness may be better understood as a reaction to deviations from highly familiar or specialized categories rather than being a response to stimuli deviating specifically on the perception of humanness.

Processing Fluency and Uncanniness

Previous researchers have suggested that the uncanniness of humanlike entities is elicited by processing disfluency caused by the entity's categorical ambiguity (e.g., Yamada et al., 2013). Ambiguity has been shown to lead to negative evaluation in faces (Halberstadt &

Winkielman, 2014). However, the present results cannot support the notion that ambiguity, or conceptual disfluency, elicits uncanniness.

The role of categorical ambiguity in the uncanny valley has been a topic of debate. Some researchers failed to show that the most ambiguous stimuli were the most uncanny (Mathur et al., 2020). Similarly, certain stimulus categories that do not straddle categorical boundaries, like faces of people with disabilities, are still rated as uncanny (Diel & MacDorman, 2021). The uncanniness of some ambiguous stimuli may also be due to those stimuli deviating from the typical configuration, which is more likely when the stimuli are straddling categorical boundaries and thus are distant from the typical. Stimuli in between two categories may be compared with both categories' typical members, leading to an increased detection of deviations. The results are in accordance with previous research showing that processing disfluency affects liking more if it elicited on a perceptual, rather than a conceptual or semantic, level (Vogel et al., 2020). As with previous research, this effect is more pronounced for configural information in more familiar categories (Chapters 2 to 4). In sum, this study provides further evidence against the effect of ambiguity on uncanniness in favour of perceptual disfluency, especially disfluency caused by deviation from specialized categories.

Deviation From Familiarity and Uncanniness

Across tasks, configural deviation of words and sentences increased uncanniness. Furthermore, the effect of deviation on uncanniness increased with language familiarity. As sufficient experience with a written language allows holistic processing of words (Björnström et al., 2014; Wong et al., 2010) and sensitivity to configural distortions (Wong et al., 2019), the moderating effect of familiarity on uncanniness can be explained by an intrinsic negative evaluation of stimuli that deviate from learned configural patterns. Familiarity has been shown to moderate the effect of configural deviation (Chapters 2 to 4). Here, the effect is

replicated with text stimuli. The results nicely fit previous suggestions that the detection of errors through the processing of high-expertise categories underlies the uncanny valley effect of near humanlike entities, especially faces (Diel & MacDorman, 2021; MacDorman & Chattopadhyay, 2016; MacDorman et al., 2009; Matsuda et al., 2012). Previously, researchers suggested an evolutionary bias to avoid oddities and anomalies in conspecifics, especially in the face (MacDorman & Ishiguro, 2006), which would not be able to explain the uncanniness of deviating written text stimuli. However, as the processing of written text may use brain areas that would otherwise be used for processing of other specialized categories (Dehaene-Lambertz et al., 2018), the negative evaluation of configurally deviating faces may also spill over to written text processing or be a general reaction towards deviants of specialized categories. If this were true, activation of stimulus-specific processing areas would be necessary for the aesthetic devaluation of deviating stimuli. In addition, uncanniness can be predicted by configural deviation of a variety of specialized categories, including voices, places and categories of trained expertise (Gauthier et al., 2006; Tanaka & Gauthier, 1997).

However, it is unclear whether deviations in general lead to aesthetic devaluation (e.g., uncanniness), or whether the subjective reaction is relative to the category's valence. Vogel et al. (2021) found that deviations from categories eliciting negative valence are experienced more positive than typical category members. Hence, deviation could actually improve aesthetic appeal of stimuli if applied to negatively perceived categories. In this sense, negative evaluation of stimuli typically associated with the uncanny valley effect may be due to the deviation from otherwise positive categories (human beings, animals, or familiar words), rather than due to deviation in itself.

In summary, Chapter 5 found evidence that the link between specialization and distortion sensitivity can be extended onto inanimate categories, namely written text. However, while uncanniness effects were found, it is unclear whether they are analogous to a proper

“uncanny valley” effect. Chapter 6 will hence present a replication of an uncanny valley function caused by deviations in another inanimate category: physical places.

Chapter 6: Structural deviations drive an uncanny valley of physical places

Methods, experiments, and large portions of the introduction and discussion in this chapter have been published in the *Journal of Environmental Psychology* (Diel & Lewis, 2022d).

Introduction

The previous chapter investigated uncanniness effects in written text. This chapter extends research of uncanniness in inanimate object categories by presenting research on an uncanny valley in physical places.

Uncanniness in physical places

Some built environments, like abandoned buildings, can elicit feelings of horror, dread, or creepiness (McAndrew, 2020). According to Kaplan's (1987) model, a high degree of mystery (defined as hidden, but "promised" information about an environment) may be elicited by surroundings not allowing inference of sufficient information, motivating further exploration. Stamps (2007) found that dim light and visual occlusion increased the mystery of physical places, which the researcher interpreted as increased informational entropy or lack of environmental information. McAndrew (2020) argued that certain physical places can be perceived as creepy if they trigger agent detection mechanisms sensitive to indicators of the presence of harmful entities. Similarly, McAndrew and Koehnke (2016) proposed that creepiness is generally elicited by *threat ambiguity*: indicators of potential danger, independent of the stimulus' category. Furthermore, absence of light may contribute to agent detection mechanisms as darkness increases the intensity of startle responses (Grillon, Pellewoski, Merikangas, & Davies, 1997; Mühlberger, Wieser, & Pauli, 2008) and enhances detection of potential threat of ethnic outgroups (Schaller, Park, & Faulkner, 2003). Thus, lack of (visual) information about the presence of threat can increase environmental creepiness.

One source of information can be schema-based typicality (Widmayer, 2002). Built environments follow predictable patterns. Houses are expected to have roofs, doors, and windows. Rooms should have entrances connected to the floor. Furniture or other features are of certain sizes, positions, and number. Certain combinations of features are predictable, like a work desk and an office chair, while others are not expected, like a toilet in a kitchen. Thus, typical physical places seem to have predictable configural patterns and can potentially deviate from those.

Visual complexity of an environment, defined as information richness, affects likability of an environment in an inverted U-shaped manner (Güclütürk, Jacobs, & Liew, 2016; Imamoglu, 2000; Kaplan, 1987). As recognizable patterns allow the organization of information to decrease complexity (Anderson, 1991), the inability to recognize learnt patterns in structurally deviating physical places may lead to a decrease of likability due to its complexity. Similarly, inconsistent scenes are less likable (Shir, Abudarham, & Mudrik, 2021), just as built and natural environments lacking in coherence, i.e., how easily an environment can be mentally organized (Coburn et al., 2020; Vartanian, Navarrete, Palumbo, & Chatterjee, 2021; Weinberger, Christensen, Coburn, & Chatterjee, 2021). Consistent or coherent places may ease recognition of typical environmental structures, allowing the identification of the specific environment, Furthermore, personally familiar spaces and spatial configurations allow for an easier wayfinding (Hölscher & Brösamle, 2007; Iftikhar, Shah, & Luximon, 2020; Wiener, Büchner, & Hölscher, 2009), and environments deviating from typical configurations may be disliked because they are more difficult to reliably traverse. In general, inconsistent or configurally deviating environments may appear less comprehensible, predictable, safe, and generally less pleasant.

One source of such configurally deviating physical places is provided through an Internet phenomenon called *liminal spaces*: a concept of real or artificial physical places judged as ambiguous or eerie (Wikimedia, 2023).

Previous research has mostly focussed on the concept of spatial *liminality* in the context of transitional places (e.g., airports) or those allowing transformative experiences (Huang, Xiao, & Wang, 2018; Neuhofer, Egger, Yu, & Celuch, 2021; Zhang & Xu, 2019). Such definitions however would be unable to explain why many of these *liminal spaces* would elicit distinct eerie or strange experiences. Hence, the term *liminal space* will here refer exclusively to such ambiguous, distressing, or "off" physical places, distinct from other definitions of liminal spaces or liminality.

While a proper academic investigation of *liminal spaces* is yet lacking, the description of *liminal spaces* as ambiguous, strange, or eerie places fits the prediction of physical places which are eerie because they deviate from the norm. Simultaneously, this study will be the first to investigate potential causes of why those specific *liminal spaces* may appear eerie or strange. Potential explanations can be based on discussed models of environmental and perceptual theories, such as a lack of place coherence (Coburn et al., 2020) inconsistent features (Shir et al., 2021), or as deviation from familiar place configurations. All in all, *liminal spaces* will here be used for the study of the perception of uncanny or creepy deviating physical places, and the uncanny valley of architecture. Hence, such stimuli will be used in the first experiment.

Research question

The present study is the first empirical investigation focussing on the uncanniness of built environments explained by the effect of configural deviation. Using the uncanny valley paradigm, an uncanny valley curve of photos of physical places is investigated by plotting

place realism against uncanniness. Furthermore, the study's goal is to test variables that may make physical places, especially those labelled as *liminal spaces*, appear uncanny. In a second experiment, the effect of direct manipulation of a physical place's configuration on uncanniness is tested, analogous to how a disruption of face configuration creates uncanny faces (Diel & MacDorman, 2021; Chapters 2 to 4). Finally, a third experiment was conducted to test how human presence interacts with the uncanniness of normal and distorted private and public places.

Predicted influences on place aesthetics

Various previous influences on place aesthetic have been suggested, with different underlying theoretical presumptions. The following variables will be investigated in the three experiments of this work:

Deviation from typical configurations. The deviation from familiarity hypothesis predicts that stimuli deviating from expected configural patterns elicit uncanniness (Chapters 2 to 4), in this case applied to deviations in built environments. Four obvious types of configural (i.e., feature-relational) deviations are 1) changes of sizes of some features compared to others, 2) the absence of expected features, 3) placement of features in unexpected positions, and 4) excessive repetition of certain features. Places containing these features should be perceived as uncannier and more abnormal.

Disgust. Disgust has been linked with uncanniness in past research (Ho & MacDorman, 2010; MacDorman & Entezari, 2015). Furthermore, atypical food variants elicit stronger disgust reactions (Koch et al., 2021). Although disgust is generally associated with organic material, distorted places may appear more unsettling for individuals with a higher disgust sensitivity.

Ambiguity. Categorical ambiguity of a stimulus has been proposed to elicit uncanniness (Cheetham et al., 2015). As places deviating from expected configurations may be more difficult to categorize and comprehend, the lack of information available on an ambiguous place may increase a sense of uncertainty.

Lighting and occlusion. Both lack of lighting and occlusion (presence of objects blocking the view of the space) contribute to a sense of mystery understood as information entropy (Stamps, 2007). Furthermore, lack of light increases anxiety responses (Grillon, Pellowski, Merikangas, & Davis, 1997; Mühlberger et al., 2008) and may thus contribute to anxiety induced in unusual places.

Social presence. The presence (or absence) of humans may influence the effects of deviating architecture. Social presence or support can act as a buffer for fear and stress responses (DeVries, Glasper, & Detilion, 2003). Social stimuli are salient (Theeuwes & Van der Stigchel, 2006) and may distract from uncanny features, or humans unreactive to unusual surroundings may normalize the subject's reactions, for example due to conformity (Cialdini & Goldstein, 2004). Human presence may also indicate safety in an environment otherwise perceived as hostile. Finally, when human presence is expected (e.g., in a public place like a mall), human absence would be a deviation from an expected configural pattern. In that sense, human presence should increase uncanniness when the presence is not expected. These explanations are investigated later.

Experiment 7

Research question and hypotheses

Experiment 7 is designed to investigate an uncanny valley curve of real and unreal physical places, including those colloquially labelled *liminal spaces*, and whether certain environmental variables can explain the effect. Hypotheses follow.

First, plotting uncanniness against place realism should create a quadratic (*U*-shaped) or cubic (*N*-shaped) function (*uncanny valley hypothesis*) akin to previous uncanny valley research (Diel et al., 2022).

Second and third, if the uncanny valley is related to *threat avoidance* (MacDorman & Ishiguro, 2006), disgust sensitivity should predict uncanniness (*disgust hypothesis*). Disugst sensitivity was measured by the revised Disgust Scale (Haidt, McCauby, & Rozin, 1994; modified by Olatunji et al., 2007), a questionnaire used in previous research linking disgust sensitivity to uncanniness (MacDorman & Entezari, 2015).

Similarly, *ambiguity tolerance* should predict uncanniness if stimuli are uncanny because of their ambiguity (e.g., Cheetham et al., 2015; *ambiguity hypothesis*). Ambiguity tolerance was measured by the ambiguity tolerance questionnaire (MacDonald, 1970), a questionnaire developed to assess individuals' differences in reaction towards ambiguous situations.

Fourth and fifth, if *threat ambiguity* underlies the uncanniness of places, threat should predict uncanniness of physical places (McAndrew & Koehnke, 2016; *threat hypothesis*). On the other hand, abnormality should predict uncanniness ratings according to the hypothesis that deviation from familiarity underlies the effect (Chapters 3 to 4; *deviation hypothesis A*).

Sixth, as previous research shows that deviations from familiar patterns may underlie the uncanny valley, distortions of the structure of places should predict uncanniness and abnormality ratings. Specifically, the level of configural deviation (feature, displacement, lack of features, repetition of features, unusual sizes) predicts uncanniness and abnormality ratings (*deviation hypothesis B*).

Lighting (*lighting hypothesis*) and visual occlusion (*occlusion hypothesis*) can increase perception of eeriness and mystery in physical places (Stamps, 2007). Each place stimulus' lighting level has been coded as *none* (major parts of the depicted place are not visible due to

lack of light), *artificial* (the place is not obscured by lack of lighting and lit with only artificial lighting), and *natural* (the place is not obscured by lack of lighting and lit with natural lighting, or both natural or artificial lighting). Occlusion was coded by whether major parts of the depicted places were not visible due to objects or architecture blocking the view.

Finally, an explorative analysis investigates why *liminal spaces* appear uncanny or abnormal to participants by focussing on qualitative responses on the most uncanny and abnormal physical places.

Materials and methods

Participants. Participants were 104 students recruited via the Cardiff University School of Psychology's Experimental Management System (EMS) and other adults recruited via Prolific®. Participants' average age was M_{age} = 29.41, SD_{age} = 9.8, and 66.67% were female. Because the motivation of Experiment 1 was exploratory and because effect size estimation was not possible for the fitted polynomial model, selection of sample size was not based on power analysis and instead based on previous research aiming to replicate an uncanny valley function (e.g., Löffler et al., 2020; Mathur & Reichling, 2016; Pütten & Krämer, 2014). Individuals with UK residence aged 18 and above with normal or corrected vision could participate. The study was approved by the Cardiff University School of Psychology Ethics Committee in May 2021 (reference number: EC.21.04.20.6342 GA).

Materials. One hundred images of real or artificial physical places were collected from various sources on the Internet. Fifty were taken from websites dedicated to the *liminal space* phenomenon[1], labelled “liminal” spaces. Twenty-five were artificial representations of places such as architectural sketches or drawings, labelled “unreal.” Finally, a set of 25 natural photographs of places were selected, labelled “real.” Latter were randomly selected from the CNN place image database (Zhou, Lapedriza, Khosla, Oliva, & Torralba, 2018).

Fifty instead of 25 “liminal” space stimuli were selected because these places were expected to be more heterogenous in their variables compared to real or artistic renditions of typical places.

Images were coded based on the following features: feature displacement, lack of features, lighting, occlusion, repetition of features, type (e.g., hallway), and unusual sizes. Coding was based on the hypotheses. All stimuli are available at https://osf.io/d9s36/.

Two questionnaires were used. First, the ambiguity tolerance questionnaire (MacDonald, 1970), consisting of 13 items (example item: “I don't tolerate ambiguous situations well”), meant to measure of how accepting or not uncomfortable a person is concerning complex issues or situations with alternate interpretations or outcomes. Second, the disgust index (Haidt, McCauley, & Rozin, 1994; modified by Olatunji et al., 2007) consisting of 14 items (example item: “If I see someone vomit, it makes me sick in my stomach”), meant to measure the degree of disgust sensitivity. In both questionnaires, items ranged from an interval of 0 (fully disagree/not at all) to 100 (fully agree/completely).

Procedure. The experiment was conducted online on the platform *pavlovia* (https://pavlovia.org). After giving informed consent, participants completed the ambiguity tolerance and disgust index questionnaires. Then participants were presented with the rating task. One hundred stimuli were presented randomly, accompanied by the four composite rating scales presented in the following order: *not eerie/creepy/uncanny – eerie/creepy/uncanny*, *strange/weird/abnormal – not strange/weird/abnormal* (reversed), *not hostile/threatening/unsafe – hostile/threatening/unsafe*, *not real/authentic – real/authentic*, ranging in an interval from 1 to 100. The first scale was selected to represent a specific negative experience related to uncanniness, the second a sense of abnormality, and the third scale threat, all constructs that were related to the uncanny valley in previous research, while

the final variable was meant to represent the independent variable of human likeness, sometimes realism, of the uncanny valley plot (see Diel et al., 2022). Definitions of eerie ("strange in a frightening or mysterious way") and uncanny ("beyond the normal or extraordinary, strangely familiar or uncomfortably strange") were provided with the experimental instructions, as well as an explanation for the real/authentic scale ("this question refers to how realistic, or close to a real-life building or place you perceive the depicted place to be"). For the other two scales, participants' subjective understanding of the terms was of interest, thus no definitions were provided. Rating scales were presented sequentially, together with each stimulus. Participants could select any point of the scales and had an unlimited time to view the image and select their response. Single scale ratings were used for the analyses, as the calculation of indices would have needed a higher number of scales which could have overtrained the participants given the high number of stimuli.

After completing the rating, participants were again presented with the 50 *liminal space* stimuli with the question if the participants thought the depicted place was strange or eerie and if so, why. Participants could type a response and confirm by pressing any arrow key. The whole procedure lasted M = 43.69 min (SD = 26.31).

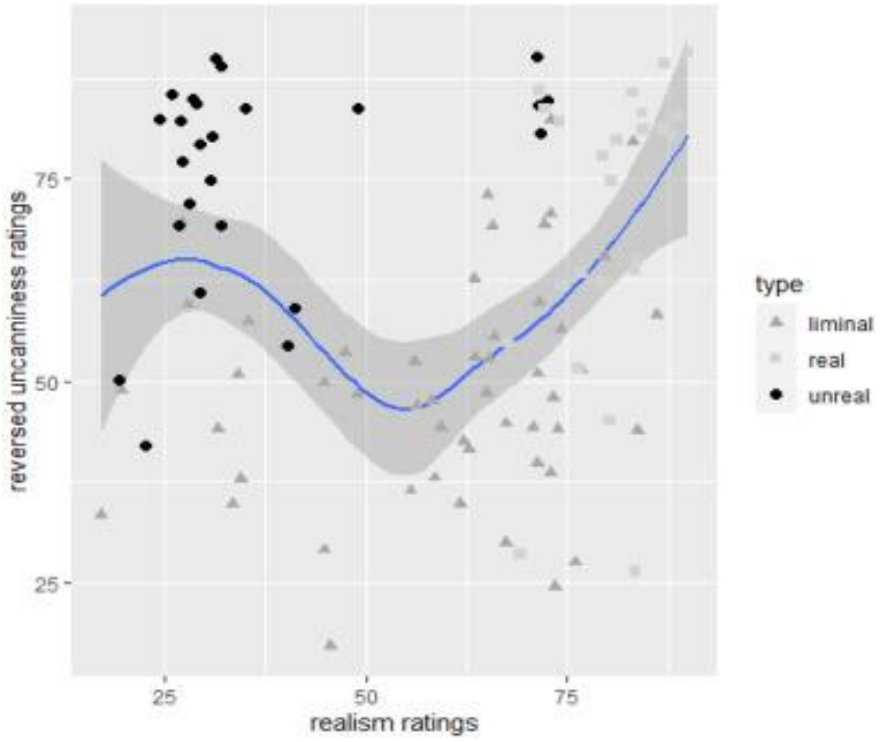
Analysis, ethics statement, and data availability. Data preparation and statistical analysis was conducted via *R*. Linear mixed models were used because they handle both fixed effects and random effects (McLean, Sanders, & Stroup, 1991), which are expected given the within-subject and within-stimulus design. This type of analysis produces the large degrees of freedom (see also Kuznetsova, Brockhoff, & Christensen, 2017; Luke, 2017). The R packages *lme4* (for linear mixed models, using the function *lmer()*) and *lmerTest* (for complete depiction of the results) were used (see Bates, Mächler, Bolker, & Walker, 2015). The data, stimuli, and R code for the analysis are available at https://osf.io/d9s36/.

Results

Uncanny valley hypothesis. Corrected coefficients of determination (R^2_{adj}) and 95% confidence intervals (CI) of regression coefficients are reported. A linear mixed model with realism (fixed effect) and stimulus and participants (random effects) was calculated to predict uncanniness. The square and cubic terms of realism were included as predictors to test the cubic and quadratic relationship akin to an uncanny valley. The results show that a linear ($t(9495) = -2.683$, $p = .007$, CI [−0.14, −0.10]), a quadratic ($t(9527) = -7.448$, $p < .001$, CI [−0.004, −0.002]), and a cubic function of realism ($t(9471) = -2.277$, $p = .023$, CI [−0.00005, −0.000004]) could all predict uncanniness. A quadratic model was a better fit than a linear model ($\chi^2 = 63.882$, $p < .001$), as was the cubic model ($\chi^2 = 69.07$, $p < .001$). The cubic was also a better fit than the quadratic model ($\chi^2 = 5.188$, $p = .022$). The adjusted coefficient of determination was $R^2_{adj} = 0.48$ for the cubic model. The data by stimulus are plotted in *Figure 6.1*. The confidence range at the highest point of uncanniness (at approx. 5 realism) falls entirely outside the confidence range for uncanniness both at lower levels of realism (e.g., 30) and higher levels of realism (e.g., 85) indicating a clear valley shape to the data. Thus, a cubic function of uncanniness and realism akin to an "uncanny valley" could best explain the data (*uncanny valley hypothesis*).

Figure 6.1

Uncanniness ratings of physical place stimuli plotted against their realism ratings, divided into the type of physical place (unreal, liminal, real). Each point in the graph corresponds to one of 100 stimuli. The line is the weighted average line of best fit and the grey shaded area is the 95% confidence range over this weighted average.



Ambiguity tolerance and disgust sensitivity. The ambiguity tolerance questionnaire's Cronbach's alpha was $\alpha = .862$ ($M = 29.86$, $SD = 10.16$), indicating good consistency. The disgust questionnaire meanwhile had a Cronbach's alpha of $\alpha = .779$ ($M = 26.91$, $SD = 17.2$), indicating acceptable consistency. Because both questionnaires showed a high correlation ($r = -0.88$), analyses were conducted independently to avoid multicollinearity.

Adjusted coefficients of determinations (R^2_{adj}) and 95% confidence intervals of regression coefficients are reported. Linear mixed model analyses with either ambiguity tolerance or disgust sensitivity (fixed effects) and stimulus and participants (random effects) showed that neither ambiguity tolerance ($t(84) = -0.807$, $p = .422$, $R^2_{adj} = 0.15$, CI [−0.13, 0.06]) nor disgust sensitivity ($t(84) = 0.02$, $p = .74$, $R^2_{adj} = 0.15$, CI [−0.08, 0.12]) predicted uncanniness. Thus, uncanniness was neither associated with disgust sensitivity (*disgust hypothesis*) nor ambiguity tolerance (*ambiguity hypothesis*).

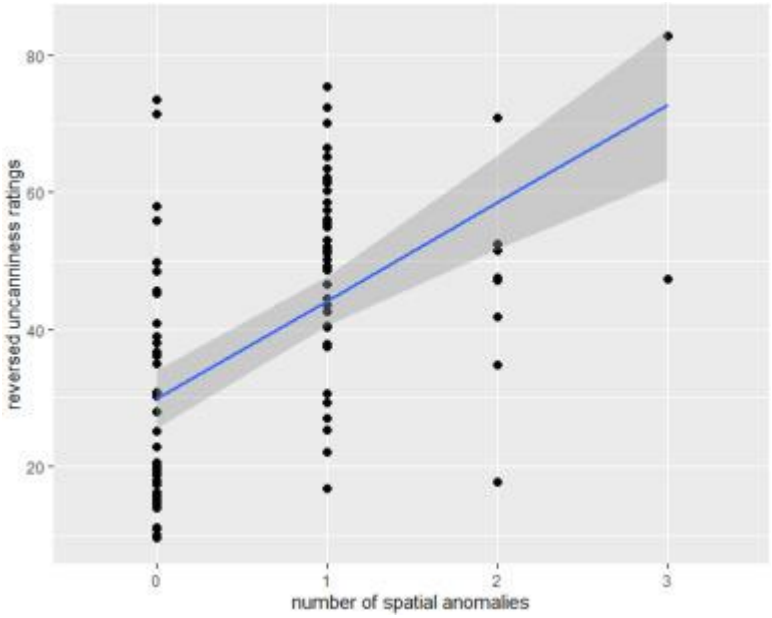
Abnormality and threat. Adjusted coefficients of determinations (R^2_{adj}) and 95% confidence intervals of regression coefficients are reported. A linear mixed model with abnormality and

threat as fixed effects and participant and stimulus as random effects showed that abnormality ($t(9353) = 27.828$, $p < .001$, CI [0.22, 0.27]), threat ($t(9599) = 33.297$, $p < .001$, CI [−0.53, 0.50]), and an interaction ($t(9576) = 4.186$, $p < .001$, CI [0.001, 0.001]) significantly predicted uncanniness. The model's determination coefficient was $R^2_{adj} = 0.58$. In total, uncanniness of physical places was associated was associated with both abnormality (*deviation hypothesis*) and threat (*threat hypothesis*).

Anomaly, lighting, and visual occlusion. Anomaly number, lighting, and occlusion have been tested as fixed effect predictors of uncanniness, and stimulus and participant as random effects. Anomaly number ($t(96) = 5.11$, $p < .001$, CI [7.48, 16.65]) and lighting ($t(96) = −2.63$, $p = .010$, CI [−13.63, −2.04]) significantly predicted uncanniness. Visual occlusion did not ($t(96) = 0.299$, $p = .766$, CI [−5.42, 7.39]). Uncanniness of different numbers of anomalies are seen in *Figure 6.2*. The determination coefficient was $R^2_{adj} = 0.48$. Thus, while both lighting type (*lighting hypothesis*) and number of anomalies (deviation hypothesis) predicted uncanniness, visual occlusion did not (occlusion hypothesis).

Figure 6.2

Uncanniness ratings of stimuli divided into their number of anomalies.



Effect of place type: hallways. Forty percent of the most uncanny stimuli in this experiment were hallway-type places, which motivated a post-hoc investigation on whether hallway-type physical places are more uncanny than other types. *T*-tests were conducted for uncanniness across all stimuli. Hallways were more uncanny than non-hallway places across all stimuli ($t(25.8) = -3.4$, $p = .001$, $d = 0.82$). Significance persisted within only *liminal space* stimuli ($t(14.99) = -1.8$, $p = .046$, $d = 0.66$). Thus, hallways are more uncanny than both typical and specifically eerie, ambiguous places.

Qualitative analysis. After excluding or shortening general responses like “it's strange” or “uncomfortable,” summarizing very similar responses (e.g., “no windows” and “windowless”), and correcting spelling errors, participant responses for the ten most uncanny stimuli are summarized in *Table 6.1*.

Table 6.1

Number of responses categorized for each content category, for both raters.

Response content	Rater 1	Rater 2
Lack of features or emptiness	49	43
Lighting or lack of lighting/darkness	39	42
Distorted sizes or proportions	30	35
Lack of safety, hostility, threat	21	20
Displacement of features	20	27
Unknown, uncertainty, lack of purpose	16	14
Repetition of patterns or features, monotony	13	16
Water	14	12
Entrapment, closed space	10	7
Dirtiness, wornness, decay	9	5
Abandonment, desolation	7	10
Visual occlusion	7	7
Lack of people	6	2

For analysis, 210 qualitative responses were categorized by content. Responses were taken of the ten most uncanny stimulus, and responses merely repeating the adjectives in the question (e.g., “the place is strange/weird/eerie/creepy”) without elaborating on the reasons were

excluded. Participants' responses were coded by two raters (authors) on whether the responses fitted one or multiple content categories via binominal yes-no responses(see Table 1), and interrater agreement was measured by calculating interclass correlations of the amounts for each content category (ICC = 0.985). Content categories were selected before coding, based on participants' responses. Data is summarized in Table 1. In total, a place's uncanniness has been most often attributed to indicators of spatial deviation like a lack of features or emptiness, distorted sizes or proportions, feature displacement, and repetition of features or patterns. In addition, uncanniness has been most often attributed to lighting or lack thereof, lack of safety, hostility, or threat, and unknown, uncertainty, or a lack of purpose. Visual occlusion or lack of people was mentioned relatively rarely.

Discussion

Results show that uncanniness plotted against realism creates a cubic function equivalent to an uncanny valley curve. Thus, the generality of the uncanny valley encompasses built environments. Furthermore, uncanniness could be predicted by both threat and abnormality and the number of anomalies. Abnormality and threat interacted, however not in a clear pattern. In the qualitative analysis participants majorly reported structural anomalies like displaced, distorted, missing, or repeating features as the sources of uncanniness. This indicates that uncanniness is driven by deviations from typical built structure. Similar to Stamps' (2007) findings, lighting predicted the uncanniness of places and was a cause of uncanniness according to qualitative ratings; visual occlusion, however, was not associated with uncanniness in the quantitative and qualitative analyses. The difference may be due to the binominal coding of visual occlusion in this study, or discrepancies in the understanding of mystery and uncanniness. Finally, neither ambiguity tolerance nor disgust sensitivity predicted uncanniness, showing that the uncanny valley of physical places is not associated with a place's ambiguity or a sense of disgust. However, it is yet unclear whether spatial

distortions elicit uncanniness or whether these variables were merely correlated in the pre-selected stimuli.

Experiment 8

Despite interesting results on the effect of deviation on uncanniness in Experiment 7, the interpretation of the results is hindered by the unstructured collection of stimulus material and the heterogeneity of places depicted. *Liminal space* stimuli are heterogenous and vary in multiple different variables, hence the causal link between structural distortion and eeriness remains unclear. Qualitative responses however indicate that structural anomalies (specifically distorted size/proportion, lack of features, displacements, and repetition) increase uncanniness of built environments. In addition, the effect of social presence was investigated, as the presence of humans can buffer fear and stress responses (DeVries et al., 2003), and social absence make a place appear more unusual when other humans are expected (e.g., public places like malls, offices, or restaurants).

Thus, a second experiment was conducted to test the effect between manipulation of configural deviation and uncanniness of built environments.

Hypotheses

To further explore the deviation from familiarity prediction that configural anomalies elicit uncanniness, the effect of spatial anomalies on uncanniness were investigated. Based on the findings in Experiment 1 that four kinds of structural anomalies were predominantly reported by participants (distorted size, lack, displacement, repetition), the following hypotheses were tested:

First, presence of spatial anomalies in a room increases uncanniness ratings.

Second, uncanniness increases with the number of distortions in a room.

Finally, the effect of social presence manipulation is investigated as social presence may have buffering effects on fear or stress (DeVries, Glasper, & Detillion, 2003). Thus, social absence should increase the uncanniness of built interiors compared to social presence.

Methods

Participants. A total of 52 participants were recruited via Prolific®. Participants' average age was M_{age} = 28.89, SD_{age} = 7.44, and 73.21% were female. Given a typical small effect size of d = 0.25 and a 2 × 6 within-subject design with five stimuli per condition (see below), a sample size of n = 52 would move the power up over .8 ($1 - \beta = 0.812$). Because no previous research on the effect of deviation on place evaluation exists, a standard small effect size was chosen (Cohen, 1988, 1992; see also Albers & Lakens, 2018; Perugini, Galluci, & Constantini, 2014) to reduce the chance of a false negative for small effects. Individuals with UK residence aged 18 and above with normal or corrected vision could participate.

Stimuli. Seventy-five images of virtual physical places were used. Stimuli were created using Roomstyler®. Five pairs of either typical or distorted versions of the same rooms were created for each of the following manipulations: *Lack* (either typical rooms or rooms lacking expected or essential features like specific furniture, doors/windows, or completely empty rooms), *repetition* (either typical rooms or rooms where certain features like furniture, patterns of furniture, doors/windows are excessively repeated to the point of being unusual or unexpected for a typical room), *displacement* (either typical rooms or rooms where certain features like furniture or doors/windows are placed in unusual or unexpected positions), and *size* (either typical rooms or rooms where certain features like furniture, doors/windows, or walls have been distorted to unusual or unexpected sizes). Because such manipulations could lead to various changes of informational value in a room, pairs of either typical or rooms with controlled distortions were created with other potential variables (lighting, escape routes, visual occlusion, visual information density) being controlled to

control *configural* room distortion specifically. The effect of the presence of humans (*social presence*) in big, open places has been investigated by creating 2 × 5 stimuli depicting either big spaces filled with human models, or without them. Human models were placed in the first and/or second plane, depending on image. One stimulus pair per distortion type, including *social presence*, is depicted in *Figure 6.3*. Stimulus design (feature manipulation) check was done by a-priori consideration. Agreement between two raters (authors) on condition-based stimulus categorization (correct nominal assignment of the stimuli to each of the 10 (2 × 5) conditions) was $\kappa = 0.74$, indicating moderate agreement.

Figure 6.3

Example stimulus pairs per distortion type. Lack = lacking features or furniture. Repetition = repeating patterns of features or furniture. Placement = displaced features or furniture. Size distortion = distorted sizes of features or furniture. Controlled distortion = distortion with other variables (lighting, escape routes, cleanness/hygiene, visual information density) controlled. Social presence = presence of human models in big, open places.


Typical
Distorted
Lack
Repetition
Placement
Size distortion
Controlled distortion
Social presence

Finally, to investigate whether increasing deviation also increases uncanniness, 3✕5 *hybrid* stimuli with either 2, 3, or 4 combined distortion types were created. Examples of hybrid stimuli including descriptions of distortions are depicted in *Figure 6.4*.

Figure 6.4

One hybrid stimulus example per number of distortion types. Types of distortions are listed below the images.

Measures and procedure. Participants rated each stimulus on the scales *not eerie – eerie*, *creepy – not creepy*, *not uncanny – uncanny*, *strange – not strange*, and *not weird – weird*, ranging in an interval from 1 to 100, allowing participants to select any point on the scales.

The experiment was conducted online. After giving informed consent, participants rated all stimuli based on the rating scales mentioned above. Stimuli were presented in a random order and simultaneously with each scale which were presented sequentially. Participants had unlimited time to view the images and select their response. The procedure lasted for about 20 min. The scales *creepy* and *strange* were reversed. The procedure took M = 27.69 min (SD = 12.04).

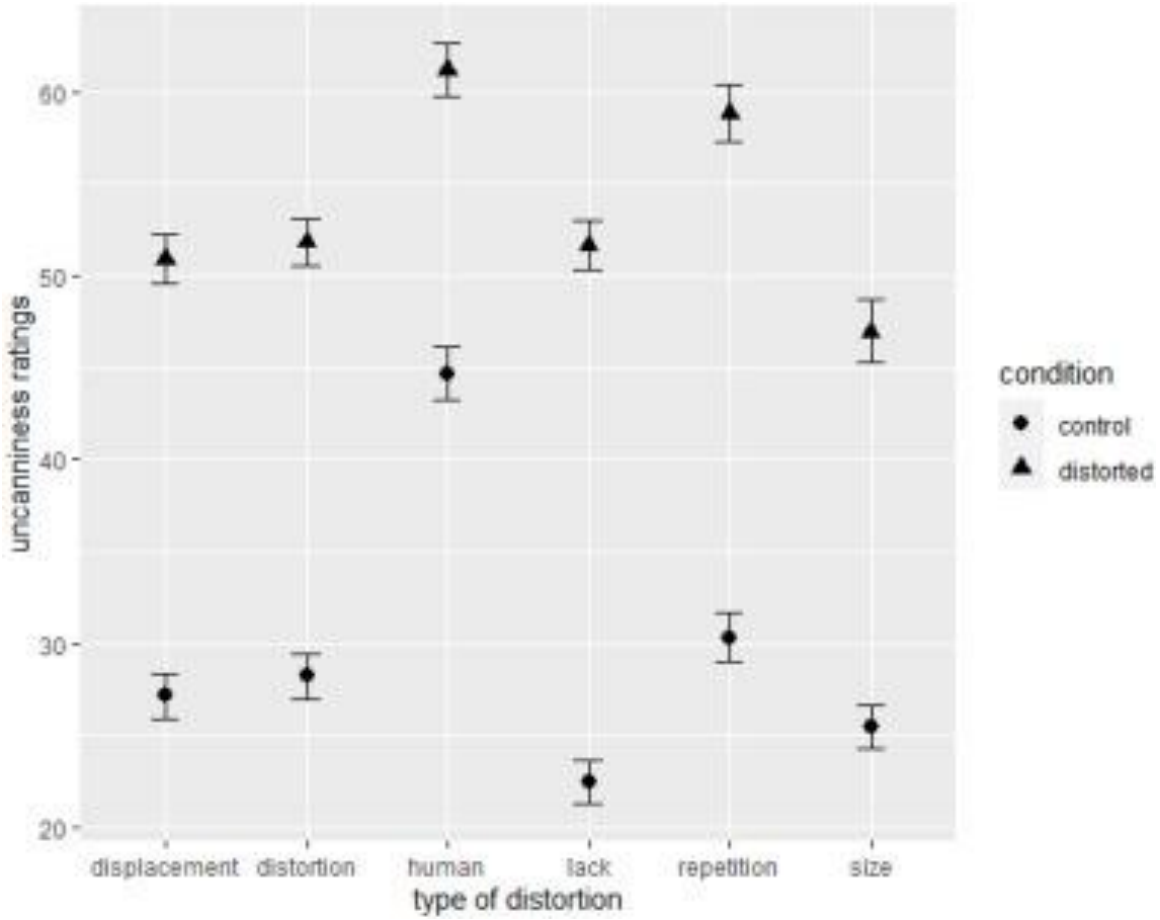
Results

Uncanniness ratings. Rating scales were combined into an *uncanniness* index by calculating the means of the five scales. The index' Cronbach's alpha was $\alpha = .89$, indicating good reliability.

Because effects of the base room pair on uncanniness were expected, linear mixed models were conducted to test the effect of distortion (control vs distortion) on uncanniness ratings for each type of distortion, with participants and base rooms as within-subject variables. Main effects of distortion for the *lack* ($t(428) = 17.728$, $p < .001$, $R^2_{adj} = 0.58$, CI [25.67, 32.06]), *repetition* ($t(425) = 16.705$, $p < .001$, $R^2_{adj} = 0.53$, CI [25.10, 32.27]), *displacement* ($t(433) = 13.44$, $p < .001$, $R^2_{adj} = 0.45$, CI [20.15, 27.04]), *size distortion* ($t(434) = 12.691$, $p < .001$, $R^2_{adj} = 0.57$, CI [20.15, 27.04]). *controlled distortion* ($t(426) = 13.87$, $p < .001$, $R^2_{adj} = 0.50$, CI [20.31, 27.01]), and *social presence* ($t(431) = 10.572$, $p < .001$, $R^2_{adj} = 0.57$, CI [13.63, 19.85]) conditions were found. Thus, all types of distortion increased uncanniness, as well as social absence. Results are summarized in *Figure 6.5*.

Figure 6.5

Uncanniness ratings for each type of distortion. Error bars depict standard errors.

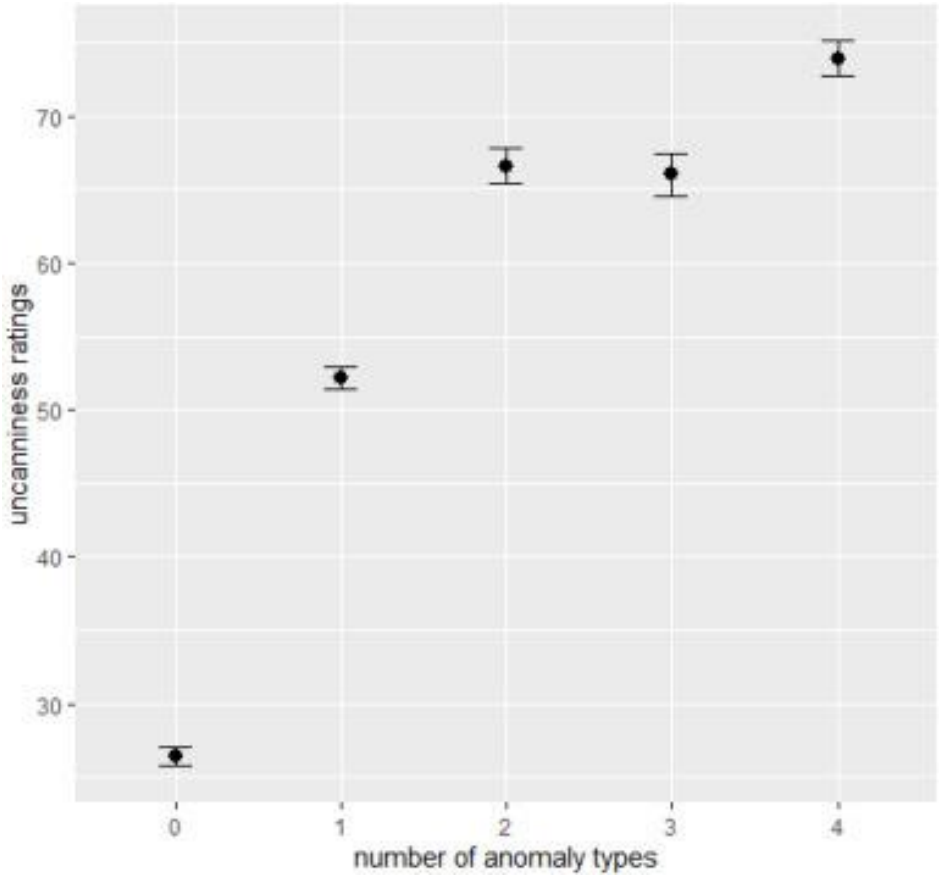


Note. Displacement = features or furniture not fitting the type of room; *distortion* = controlled configural distortion; *human* = presence (control) or absence (distorted) of humans in spatially distorted places; *lack* = lack of features or furniture; *repetition* = excessively repetitive features or furniture; *size* = unusual or distorted sizes.

Finally, number of anomalies predicted uncanniness ($t(2616) = 38.28$, $p < .001$, $R^2_{corr} = 0.47$, CI [11.86, 13.14]), summarized in *Figure 6.6*. Thus, uncanniness of a physical place increases with the number of structural anomalies present.

Figure 6.6

Uncanniness ratings averaged across number of anomaly types in a room. Error bars depict standard errors.



Discussion

Experiment 8 showed that the uncanniness of physical places is driven by the presence and number of spatial anomalies like lack of, repeating, displaced, or proportionally distorted features. Effects were present even if other variables were controlled, indicating that configural distortion of a place elicits uncanniness. Finally, social presence in places decreases uncanniness. However, the reason behind the effect of social presence is unclear and other objects could potentially play the role as other humans. This is explored further in Experiment 9.

Experiment 9

Experiment 8 showed that social presence decreases uncanniness of big, open interiors. Multiple explanations are possible: that humans act as distractors from unusual features (*distraction*), that a lack of humans does not fit the configuration of wide places (*deviation*),

that humans normalize the oddity of a distorted physical place (*normalization*), and that social presence decreases potential threat (*threat*). Finally, results from Experiment 7 and 8 both indicate that emptiness can increase uncanniness of physical places, the social presence stimuli in Experiment 8 did not control for physical emptiness. Thus, a Experiment 9 has been conducted to test the hypotheses mentioned above.

Hypotheses

According to the *distraction hypothesis*, human presence decreases uncanniness of a place because social stimuli distract from spatial anomalies due to their salience. As a short display of a stimulus would shift the attentional bottleneck towards more salient stimuli (Itti, 2005; Theeuwes & Van der Stigchel, 2006), social presence should decrease the viewers' ability to detect spatial anomalies in quickly displayed stimuli. Thus, when a place is briefly presented (500 ms), participants should be less able to detect architectural anomalies or oddities when humans are present in the image, regardless of whether the place is private or public (*distraction hypothesis*).

According to the *deviation hypothesis*, social presence would decrease uncanniness of places when humans are expected in those places Diel & MacDorman, 2021; Shir et al., 2021), like malls, restaurants, or busy streets. If deviation from expectation would predict uncanniness, the presence of humans would however also *increase* uncanniness of places where presence is unexpected, such as toilets or bedrooms. Thus, and interaction between the type of place (social presence expected vs unexpected) and social presence is expected. As social presence would generally be expected in public places and unexpected in private places, human presence should decrease uncanniness of public places (malls, fitness studios, offices, etc.) and increase the uncanniness of private places (e.g., home rooms; *deviation hypothesis*).

The *normalization hypothesis* predicts that social presence normalizes abnormality and thus uncanniness in general, for example due to the calm and friendly demeanours of human models that could elicit similar reactions in viewers through conformity (Cialdini & Goldstein, 2004). Social presence should thus decrease abnormality and uncanniness of distorted places, regardless of whether the place is private or public (*normalization hypothesis*).

Finally, the *threat hypothesis*, built upon the *threat ambiguity* (McAndrew & Koehnke, 2016), predicts that social presence generally decreases threat as the presence of other humans decreases the chance of potential danger like hiding predators or hazards in abandoned places. Social presence should thus decrease both threat and uncanniness, regardless of whether the place is distorted, or a public or private place (*threat hypothesis*).

Methods

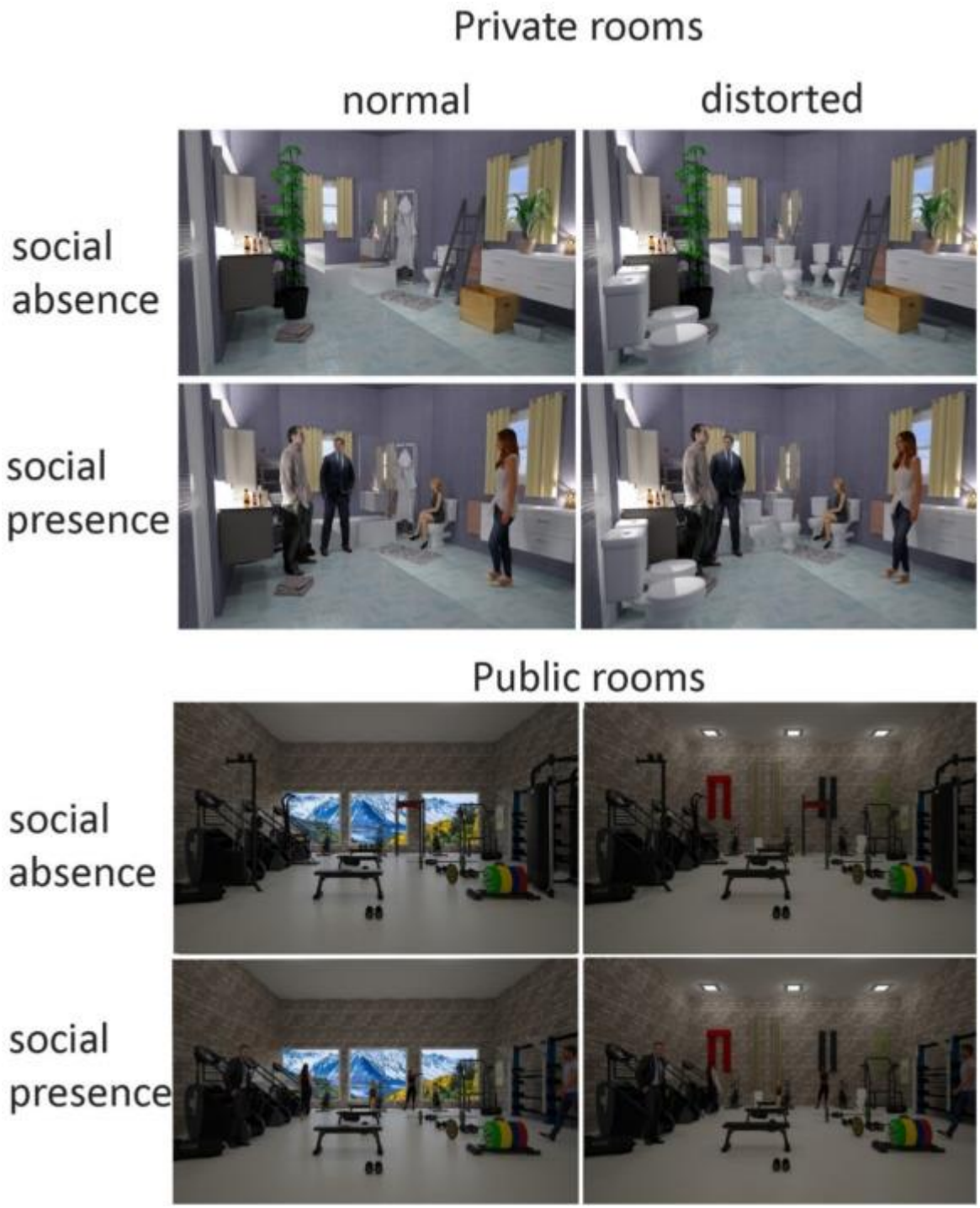
Participants. Thirty-seven participants were recruited via Prolific®. Participants' average age was M_{age} = 24.19, SD_{age} = 4.88, and 74.19% were female. Given an effect size of d = 0.25, a n = 37 sample and a 2 × 2 × 2 within-subject design with five stimuli per condition (see below), power would exceed 0.8 (1-β = 0.841). Because no previous research on the effect of deviation on place evaluation exists, and because finding the existence of an effect, even a small one, was the goal of the experiment, a standard small effect size was chosen (Cohen, 1988, 1992; see also Albers & Lakens, 2018; Perugini, Gallucci, & Costantini, 2014) to reduce the chance of a false negative for small effects. Individuals with UK residence aged 18 and above with normal or corrected vision could participate.

Stimuli. Quadruplets of rooms were created as stimuli using Roomstyler®. The same base room was used to manipulate social presence (human models or furniture) and distortion (typical rooms or distorted versions based on distortion types in Experiment 2). Finally,

rooms were either private (bathroom, kitchen, living room, bedroom, hallway) or public (fitness studio, underground hallway, office, supermarket, lecture hall). Thus, stimuli were divided based on a 2 × 2 × 2 design with social presence, distortion, and room type as independent variables, with five stimuli per condition, adding up to a total of 40 stimuli. Human models were selected matched to the place (e.g., models wearing gym clothes for a gym) and placed to indicate meaningful actions or interactions. To control for the effect of emptiness, human models were replaced with place-typical furniture of around the same size as the models. Human models were placed in the first and/or second plane, depending on image. *Figure 6.7* depicts example stimuli for each condition. Stimulus design (feature manipulation) check was done by a-priori consideration. Agreement between two raters (authors) on condition-based stimulus categorization (a series of three binominal yes-no assignments per stimulus: private/public, social presence/absence, and distorted/normal) was $\kappa = 0.83$, indicating strong agreement.

Figure 6.7

Example images for each condition. The type of distortion for both the private and public room is repetition (of toilets or windows).



Procedure. The experiment was conducted online and consisted of two parts. After giving informed consent, participants viewed each of the 40 stimuli randomly for 500 ms, preceded and followed by grey noise of 500 ms. After each stimulus, participants were asked two questions with scales ranging from *totally disagree* (1) to *totally agree* (100): “The room's architecture or design was unusual or strange.” and “I saw some oddities in the room.”

For the second part, participants again viewed all 40 images presented in a random order and were asked to rate the places on 7 scales, each ranging as intervals from (1) to (100): *not eerie – eerie*, *not creepy – creepy*, *uncanny – not uncanny*, *not strange – strange*, *not weird – weird*, *not threatening – threatening*, and *unsafe – not unsafe*. Participants were allowed to select any point on the scales. The procedure lasted for M = 18.23 min (SD = 9.20).

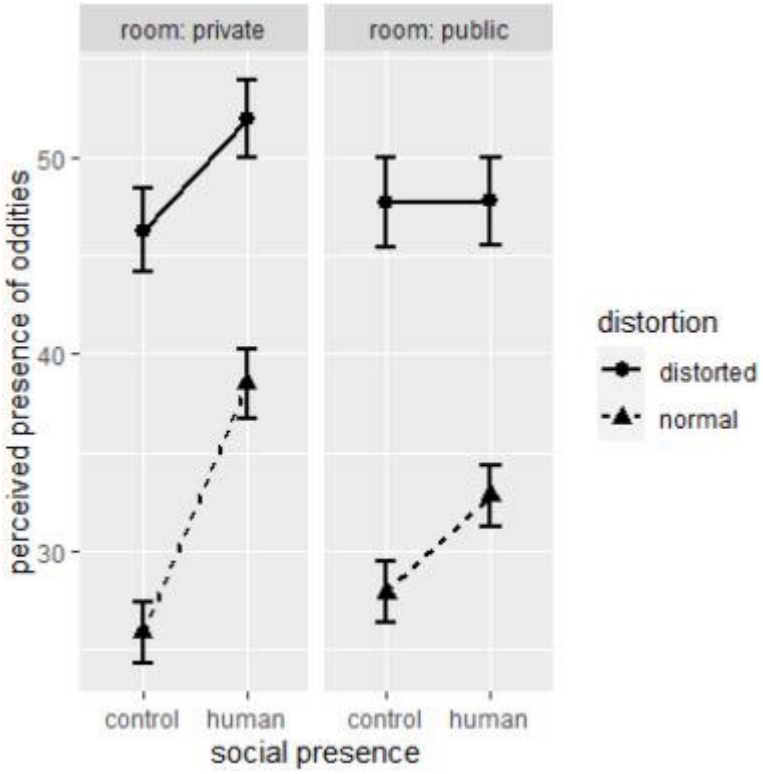
Results

Rating scales. For the first part of the experiment, the two questions on the rooms' strangeness or oddities were combined into a single *perception of oddities* variable (Cronbach's alpha α = .84). For the second part, rating scales were combined into the indices uncanny (*eerie, creepy, uncanny*), abnormal (*strange, weird*), and threatening (*threatening, unsafe*). Cronbach's alphas were α = .87, α = .97, and α = .81, respectively.

Distraction hypothesis. A within-subject ANOVA was conducted for the perception of oddities during 500 ms presentation, with social presence, distortion, and room type as within-subject variables. The data are presented in *Figure 6.8*. Results show main effects of both social presence (F(1, 36) = 12.27, p = .001, η^2_p = .25, 95% CI [0.05, 0.47]) and distortion (F(1, 36) = 161.49, p < .001, η^2_p = .82, 95% CI [0.7, 0.88]), and interaction effects between social presence and distortion (F(1, 36) = 9.1, p = .005, η^2_p = .2, 95% CI [0.02, 0.42]) and social presence and room type (F(1, 36) = 10.5, p = .003, η^2_p = .23, 95% CI [0.03, 0.44]). No other terms were significant.

Figure 6.8

Mean perceived oddities ratings across conditions. Error bars indicate standard errors.



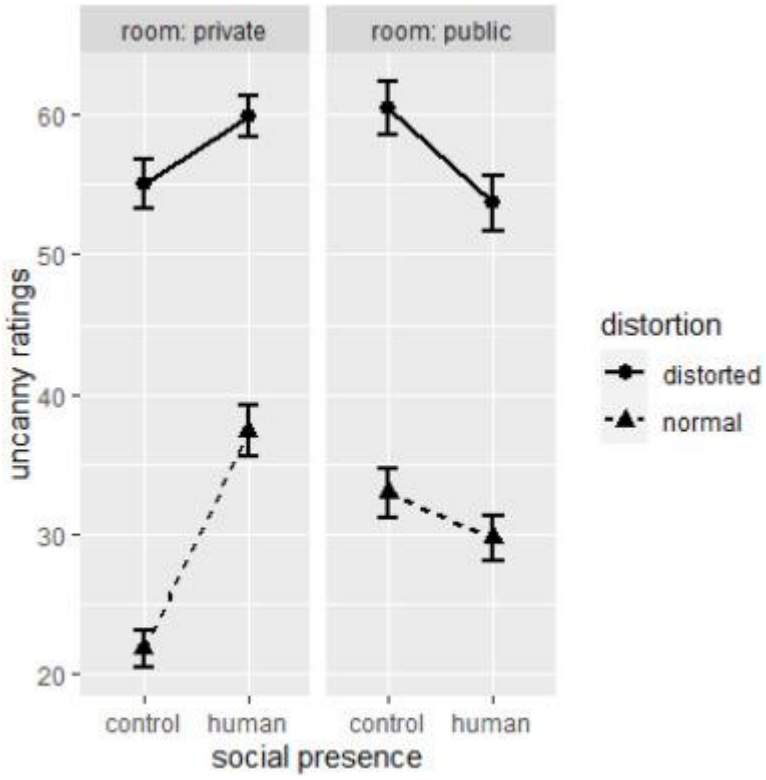
Post-hoc Tukey tests were conducted to test whether social presence decreased the detection of oddities. Results show that distortion increased perception of oddities in all social presence × room type conditions ($t(120) = 9.7$, $p_{adj} < .001$, $d = 1.77$, 95% CI [1.35, 2.19], for control private; $t(120) = 6.47$, $p_{adj} < .001$, $d = 1.18$, 95% CI [0.79, 1.57], for human private; $t(120) = 9.41$, $p_{adj} < .001$, $d = 1.72$, 95% CI [1.3, 2.13], for control public; $t(120) = 7.1$, $p_{adj} < .001$, $d = 1.3$, 95% CI [0.9, 1.69], for human public). However, social presence did not decrease perception of oddities in any condition ($t(101) = -2.49$, $p_{adj} = 1.000$, for distorted private; $t(101) = -5.51$, $p_{adj} = 1.000$, for normal private; $t(101) = 0.02$, $p_{adj} = 1.000$, for distorted public; $t(101) = -2.14$, $p_{adj} = 1.000$, for normal public). Additional exploratory post-hoc test were conducted: social presence increased oddity perception in distorted private ($t(101) = -2.49$, $p_{adj} = .029$, $d = -0.5$, 95% CI [−0.89, −0.10]) and normal private ($t(101) = -5.51$, $p_{adj} < .001$, $d = -1.1$, 95% CI [−1.51, −0.68]) places, but not in distorted public ($t(101) = 0.02$, $p_{adj} = 1.000$) or normal public places

($t(101) = -2.14$, $p_{adj} = .07$). As social presence did not decrease the detection of oddities in any place condition, hypothesis 1 was not supported.

Deviation hypothesis. A within-subject ANOVA was conducted for *uncanny* ratings, with social presence, distortion, and room type as within-subject variables. Data are summarized in *Figure 6.9*. Results show a main effect of distortion ($F(1, 36) = 179.18$, $p < .001$, $\eta^2_p = .83$, 95% CI [0.72, 0.89]), and interaction effects between social presence and distortion ($F(1, 36) = 15.82$, $p < .001$, $\eta^2_p = .31$, 95% CI [0.08, 0.51]), and between social presence and room type ($F(1, 36) = 62.61$, $p < .001$, $\eta^2_p = .63$, 95% CI [0.43, 0.76]).

Figure 6.9

Mean uncanniness ratings across conditions. Error bars indicate standard errors.



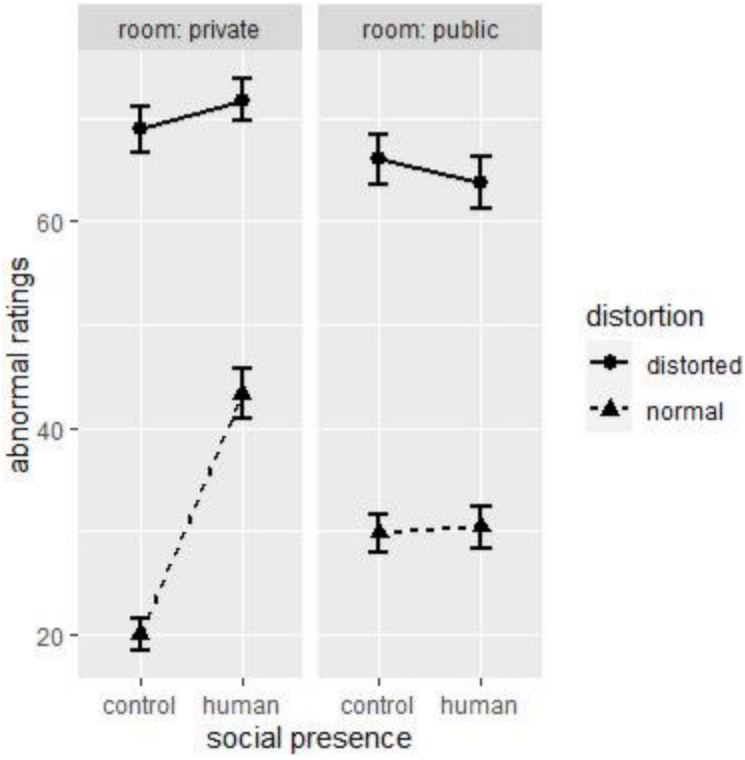
Post-hoc Tukey tests show that distortion increased uncanniness regardless of social presence or room type ($t(85) = 8.51$, $p_{adj} < .001$, $d = 1.85$, 95% CI [1.33, 2.35], for control private; $t(85) = 12.95$, $p_{adj} < .001$, $d = 2.81$, 95% CI [2.21, 3.4], for human private; $t(85) = 10.59$, p_{adj}

<.001, d = 2.3, 95% CI [1.75, 2.84], for control public; $t(85) = 9.58$, p_{adj} <.001, d = 2.08, 95% CI [1.55, 2.6], for human public). Furthermore, social presence decreased uncanniness of distorted public places ($t(133) = 2.95$, $p_{adj} = .01$, d = 0.51, 95% CI [0.17, 0.86]), but not normal public places ($t(133) = 1.72$, $p_{adj} = .265$), and increased the uncanniness of both distorted private ($t(133) = -2.04$, $p_{adj} = .043$, d = −0.35, 95% CI [−0.7, −0.01]) and normal private ($t(133) = -7.41$, p_{adj} <.001, d = −1.29, 95% CI [−1.66, −0.91]) places. As social presence increased the uncanniness of private places but decreased the uncanniness of distorted (but not normal) public places, hypothesis 2 was mostly supported.

Normalization hypothesis. A within-subject ANOVA with social presence, distortion, and room type as within-subject variables was used to investigate the effect of these variables on abnormality ratings. The data are summarized in *Figure 6.10*. Main effects were observed for social presence ($F(1, 36) = 17.16$, $p < .001$, $\eta^2_p = .32$, 95% CI [0.13, 0.5]) and distortion ($F(1, 36) = 191.71$, $p < .001$, $\eta^2_p = .84$, 95% CI [0.76, 0.89]), and interactions between social presence and distortion ($F(1, 36) = 31.22$, $p < .001$, $\eta^2_p = .46$, 95% CI [0.26, 0.61]), social presence and room type ($F(1, 36) = 33.31$, $p < .001$, $\eta^2_p = .48$, 95% CI [0.28, 0.62]), and all factors combined ($F(1, 36) = 12.04$, $p = .001$, $\eta^2_p = .25$, 95% CI [0.07, 0.43]).

Figure 6.10

Mean abnormal ratings across conditions. Error bars depict standard errors.

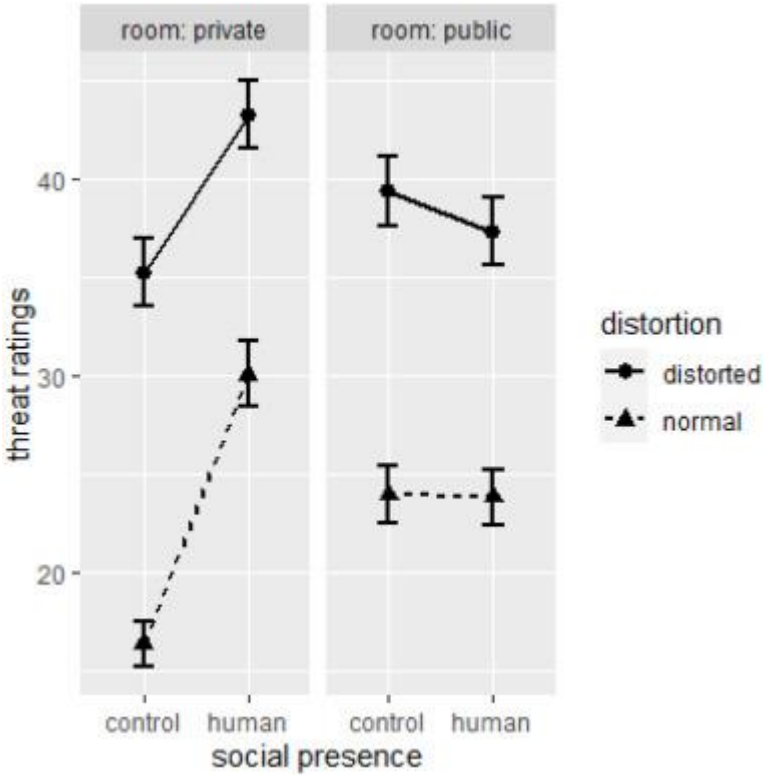


Post-hoc Tukey tests were calculated to test the specific predictions. Distortion increased abnormality in all social presence × room type conditions ($t(84) = 14.47$, $p_{adj} < .001$, $d = 3.16$, 95% CI [2.51, 3.79], for control private; $t(84) = 8.12$, $p_{adj} < .001$, $d = 1.77$, 95% CI [1.26, 2.27], for human private; $t(84) = 19.66$, $p_{adj} < .001$, $d = 4.29$, 95% CI [3.51, 5.06], for control public; $t(84) = 10.08$, $p_{adj} < .001$, $d = 2.2$, 95% CI [1.65, 2.74], for human public). Social presence however did not decrease abnormality in any distortion × room type condition ($t(135) = -0.77$, $p_{adj} = 1.000$, for distorted private; $t(135) = -9.18$, $p_{adj} = 1.000$, for normal private; $t(135) = 0.67$, $p_{adj} = 1.000$; for distorted public; $t(135) = -0.22$, $p_{adj} = 1.000$, for normal public). Additional explorative post-hoc tests revealed a significant increase of abnormality ratings when humans were present in undistorted private places ($t(135) = -9.18$, $p_{adj} < .001$, $d = -1.58$, 95% CI [−1.96, −1.19]), but not if the same place was distorted ($t(135) = -0.77$, $p_{adj} = .445$). As social presence did not decrease abnormality ratings in any condition, hypothesis 3 was not supported.

Threat hypothesis. A within-subject ANOVA has been conducted to test the effect of the within-subject variables social presence, distortion, and room type on threat ratings. Data are summarized in *Figure 6.11*. Main effects were observed for social presence (F(1, 36) = 23.44, p < .001, η^2_p = .39, 95% CI [0.19, 0.56]) and distortion (F(1, 36) = 64.08, p < .001, η^2_p = .64, 95% CI [0.47, 0.74]). and interactions between social presence and distortion (F(1, 36) = 4.16), p = .049, η^2_p = .1, 95% CI [0.00, 0.28]) and social presence and room type (F(1, 36) = 30.66, p < .001, η^2_p = .46, 95% CI [0.26, 0.61]). No other term was significant.

Figure 6.11

Mean threat ratings across conditions. Error bars depict standard errors.



Post-hoc Tukey tests again show that threat was higher for distorted compared with normal places across social presence ⨯ room type (t(95) = 7.67, padj <.001, d = 1.57, 95% CI [1.11, 2.03], for control private; t(95) = 4.98, padj <.001, d = 1.02, 95% CI [0.59, 1.45], for human

private; $t(95) = 5.98$, $p_{adj} < .001$, $d = 1.23$, 95% CI [0.69, 1.66], for control public; $t(95) = 5.35$, $p_{adj} < .001$, $d = 1.1$, 95% CI [0.66, 1.53], for human public). However, threat was not decreased by human presence in any distortion × room type condition ($t(144) = -3.45$, $p_{adj} = 1.000$, for distorted private; $t(144) = -6.8$, $p_{adj} = 1.000$, for normal private; $t(144) = 0.84$, $p_{adj} = 1.000$, for distorted public; $t(144) = 0.07$, $p_{adj} = 1.000$, for normal public). After examining the t values, additional exploratory post-hoc tests were conducted and showed that threat was increased by social presence in normal ($t(144) = -3.45$, $p_{adj} < .001$, $d = -0.58$, 95% CI [−0.91, −0.24]) and distorted ($t(144) = -6.8$, $p_{adj} < .001$, $d = -1.13$, 95% CI [−1.13, −0.78]) private places. As social presence did not decrease threat ratings, hypothesis 4 was not supported.

Discussion

Experiment 9 investigated explanations on the effect of social presence on environmental uncanniness. Humans neither distracted from, nor normalized spatial anomalies. However, social presence either increased or decreased the uncanniness, oddity, abnormality, and threat partially depending on whether humans would be expected. As human models were replaced with furniture in the social absence conditions, changes of uncanniness are likely not due manipulation of physical emptiness itself. Thus, the effect of social presence depends on whether humans are expected or not, fitting the deviation from familiarity prediction.

General discussion

This study was motivated to investigate the effect of deviation from familiarity on the evaluation of built environments. Specifically, it was the first to investigate whether an uncanny valley can be found for physical places, and whether uncanniness of places can be explained by configural deviations. Results and their implications for the uncanny valley effect and the evaluation of built environments are discussed.

Discussion of results

An uncanny valley of physical places. Experiment 7 found that a cubic function of realism can best explain uncanniness for naturalistic images of physical places, comparable to the uncanny valley typically observed in uncanny valley research (Mori, 2012): Physical places become more likable with increasing realism, but deviations from typical structural patterns of realistic places are rated strange or eerie.

Most stimuli within the "valley"-range of the function are *liminal space* type places while those left to the valley are unreal places. This pattern follows results of previous research on the uncanny valley: unrealistically human, mechanical robotic entities lie to the left of the valley while uncanny stimuli are characterized by highly realistic yet "off" exemplars, for example due to atypical or mismatching features (Mathur & Reichling, 2016; Mori, 2012). The eeriness or strangeness of *liminal spaces* similarly can be explained by their deviation from otherwise typical and realistic physical places.

It is not clear whether the uncanniness observed here equates the uncanniness typical for the uncanny valley, as uncanniness can be elicited by various stimuli and situations. However, the cubic *N*-shaped function and the effect of deviation from familiar patterns found here are characteristic to previous uncanny valley research (Diel & MacDorman, 2021; Mori, 2012). Similar statistical patterns indicate that the mechanisms underlying the uncanny valley of physical places are comparable to those observed in uncanny valley research. The previous emphasis on humanoid stimuli in uncanny valley research may reflect humans' high perceptual familiarity and a narrow range acceptable of human appearance, causing even slight deviations of manufactured androids to be uncanny, while such deviations would typically not occur when constructing built environments.

However, this work shows that built places deviating from typical configurations also elicit uncanniness. Thus, the uncanny valley observed for both places and biological stimuli may have the same underlying cognitive mechanisms not bound by stimulus category. The present results support the notion that the uncanny valley is not restricted to human or animal stimuli (which is assumed in some theories like disease avoidance or dehumanization), and explanations of the phenomenon should be applicable independent of stimulus categories, examples including categorization-related processes (Cheetham et al., 2015), deviation from familiarity (Chapters 2 to 4), expectation violation (MacDorman & Ishiguro, 2006), and threat ambiguity (McAndrew & Koehnke, 2016).

Uncanniness and configural deviation. Configural deviation is a potential source of environmental uncanniness. In all experiments, abnormality, structural anomalies, or deviations from typical built environments (including the expected presence or absence of people) were associated with uncanniness. The results align with previous research finding that configural deviations in faces are uncanny (Chapters 2 to 4; Diel & MacDorman, 2021; Mäkäräinen, Kätsyri, & Takala, 2014), and that inconsistent features in scenes are weird, disturbing, and less likable (Shir et al., 2021). Uncanniness could thus result from deviations from familiar configurations.

Previous research found associations between reduced environmental likability and a lack of coherence (Coburn et al., 2020; Kaplan, 1987; Vartanian et al., 2021; Weinberger et al., 2021). Configural deviation could decrease a place's perceived coherence understood as a disagreement between a place's elements and in turn likability.

Social presence decreased uncanniness ratings of wide, deviating places in Experiments 8 and 9. The effect of human presence on uncanniness was however not general and interacted with the type of rooms: Human presence decreased uncanniness in wide places, yet increased

uncanniness in private rooms. Humans neither distract from, nor normalize spatial anomalies. Furthermore, the effect of physical emptiness observed in Experiment 2 was controlled as human models were replaced with furniture in the social absence conditions. Instead, the uncanniness-decreasing effect of social presence on distorted public places may reflect eased recognition of a place based on typicality, making it less deviating, while increasing deviation (and threat) of private places. Alternatively, human presence may decrease an ambiguous or unfamiliar place's threat since a threatening place is less likely to be inhabited.

If the deviation from familiarity explanation were correct, the number of human models present should further moderate the effect of social presence on uncanniness depending on place: A fewer humans (e.g., one or two) may be acceptable in some private places like living rooms or kitchens, however social presence should become unacceptable when a certain threshold is reached. The effect may be further moderated by the familiarity of individuals (and places) depicted, as seeing a familiar person in an unusual location, or an unfamiliar person in a personal location may further estrange a scene. Future research can look into how the number and familiarity of people influences uncanniness ratings of places.

Threat and lack of information. Threat significantly predicted uncanniness in Experiment 7. Participants furthermore reported lighting, lack of safety and threat, and visual occlusions as reasons for a place's eeriness. However, only lighting, not occlusion, significantly predicted uncanniness. Lack of light has been associated with perceived lack of safety in past research (Boomsma & Steg, 2014). These results align with McAndrew and Koehnke's (2016) theory of *threat ambiguity* and with Stamps (2007) observations that lighting and occlusion increase a place's sense of mystery or lack of information.

While threat ambiguity can explain the perception of threat of uncanny places, it is unclear whether the *ambiguity* of threat elicited eeriness (as proposed by McAndrew & Koehnke,

2016). Feelings of threat may have stemmed from other sources: detecting potential environmental hazards, uncleanness, and other sources of contaminations in relation to threat avoidance theories (MacDorman & Ishiguro, 2006). Alternatively, stimuli deviating from familiar patterns may be threatening because they do not fit established cognitive conceptualizations or categories (Mangan, 2015; Schoenherr & Burleigh, 2015) and are thus less predictable. Thus, while uncanniness ratings correlated with threat, it is still unclear whether threat ambiguity specifically causes uncanniness of deviating architecture.

Recognizable patterns and structures allow to infer category-based information (Widmayer, 2002): Recognizing a place as a private bathroom provides additional relevant information. Anomalous or pattern-deviating places however escape categorization and prohibit inference of useful information: as a result, such places may appear eerie, strange, less safe, and potentially mysterious (Stamps, 2007).

Because environmental safety is of value to residents and a lack of perceived safety is related to stress and poor mental and physical health (Bilotta, Ariccio, Leone, & Bonaiuto, 2019; Brosschot, Verkuil, & Thayer, 2018; Conde & Pina, 2014), Designing environments based on typicality and predictability can increase residents' comfort and protect their health.

Configural deviation and the aesthetics of physical places

The present research shows that built environments can cause a sense of eeriness or uncanniness if they sufficiently deviate from familiar, expected patterns and structures. These results can provide insights into understanding a variety of research on the evaluation of built environments:

Bizarre or postmodern architecture has been described as not fitting typical categories of places and structures (Jencks, 1979), and buildings of such styles are judged as less typical,

familiar, pleasant, and preferable (Purcell, 1995; Purcell & Nasar, 1992; Stamps & Nasar, 1997). Thus, their deviation may decrease their likability.

Previous research found that a lack of coherence may reduce the likability of built and natural environments, potentially by increasing cognitive disfluency (Coburn et al., 2020; Vartanian et al., 2021; Weinberger et al., 2021). Highly incoherent places lacking internal organization may fall in the uncanny valley of architecture observed in Experiment 1.

Images typically described as *liminal spaces* in Internet communities may appear eerie, strange, or uncanny because the depicted places deviate from typical, experience-based expectations of places, and could thus be considered place-analogies of the uncanny valley. *Liminal space* stimuli, their place typicality, and aesthetic appeal could be investigated in future research, for example in the context of coherence (Kaplan, 1987) or expected pathfinding ability related to place familiarity (Hölscher & Brösamle, 2007), but also potential positive ratings of deviating or "liminal" spaces. Finally, recognizing a place can support navigation (e.g., by inferring relevant information), and familiarity helps with wayfinding (Haq & Zimring, 2003; Hölscher & Brösamle, 2007). Perceived difficulties in wayfinding and walkability are associated with increased anxiety (Chang, 2013) and decreased likability (Li, 2006), and well-being (Jaskiewicz & Besta, 2014). As a deviating place configuration may reduce the amount of information one may infer about the environment and its navigation, it may also negatively affect likability and well-being. In summary, negative reactions towards distorted, changed, or otherwise unexpected built environments are found throughout literature. A deviation-from-familiarity framework can encompass these reactions towards deviations from structural patterns of places, changes in specific environments, as well as the eerie atmosphere of configurally disordered or anomalous places, such as those observed in "haunted" settings. The aesthetics of built

environments could thus be improved by designing them to adhere to their expected typicality.

Chapters 2 to 4 presented evidence for a statistical link between specialization and distortion sensitivity, while Chapters 5 and 6 applied this link onto inanimate categories. Such a link could explain the uncanny valley and uncanniness effects in general. Yet multiple theories and explanations on the uncanny valley have been proposed over the past decades, with little to no critical investigation. The following chapters will thus focus on testing hypotheses of the uncanny valley's theories against each other, starting with an uncanny valley of voice stimuli in Chapter 7.

Chapter 7: The vocal uncanny valley: Deviation from typical organic voices best explains uncanniness

Methods, experiments, and large portions of the introduction and discussion in this chapter are currently in review in the journal *Scientific Reports*.

Introduction

The following chapters present results critically investigating theories of the uncanny valley. In this chapter, the refined theory is contrasted with categorization- and mind attribution-based accounts of the uncanny valley, while an uncanny valley is also replicated using voice stimuli.

The vocal uncanny valley

The refined moderated model of the uncanny presented here is domain-general and thus should occur in auditory stimuli like voices as well. An uncanniness effects have been observed in the context of android appearance and behaviour and their mismatch with voices (Meah & Moore, 2012; Mitchell et al., 2011). However, previous research has consistently failed to find a 'vocal uncanny valley' when isolated voice stimuli were used: likability increased with a voice's human likeness (Baird et al., 2018a, 2018b; Kimura et al., 2018; Kühne et al., 2020; Romportl, 2014). However, except for one study (Kimura et al., 2018), all researchers investigating a vocal uncanny valley have used exclusively synthetic voices and/or fully human voice stimuli. There are four explanations on why an uncanny valley of voices may not have been found: 1) a vocal uncanny valley does not exist; 2) stimulus selection has sufficient range but lacks stimuli that fall into the valley; 3) stimulus selection does not extent into the valley and stops before the drop (Mara et al., 2022); 4) stimulus selection begins at the valley and ends at full human likeness. These explanations urge different implications for the design of artificial voices: If an uncanny valley of voices has not yet been reached, technological development may yet lead to its emergence. If, on the

other hand, today's synthetic voices already overcome an uncanny valley or if a vocal uncanny valley does not exist, then this particular issue can be disregarded for the design of artificial voices.

Deviation and typicality in voices

Analogous to faces, voices can be defined based on their typicality. Certain disorders related to the vocal tract, like vocal fold paresis, Reinke's Edema, or muscle tension dysphonia, can lead to changes in the voice. Pathological voices are more likely to be categorized as atypical (Kreiman et al., 2018; Kreiman & Gerratt, 2003; Kreiman et al., 1992) and are evaluated more negatively across various social dimensions compared to healthy voices (Altenberg & Ferrand, 2006; Amir & Levine-Yundof, 2013; Eadie et al., 2017; Schroeder et al., 2020). In analogy, previous research has suggested that dysmorphic, diseased, or very unattractive faces are perceived as uncanny or creepy (Corradi et al., 2021; Diel & MacDorman, 2021). Thus, pathological voices, similarly to disfigured faces, may fall into an uncanny valley as highly realistic yet deviating stimuli.

Uncanniness and categorization difficulty

Stimuli difficult to categorize may fall into an uncanny valley (Chattopadhyay & MacDorman, 2016; Cheetham et al., 2013; Yamada et al., 2013. Categorization difficulty may decrease likability due to processing disfluency (Carr et al., 2017; Winkielman et al., 2003) or cognitive conflict (Weis & Wiese, 2017). As categorization theories do not depend on stimulus domain, categorical ambiguity should thus also predict the uncanniness of voices.

Experiment 10

The aim of the experiment is to investigate the existence of a vocal uncanny valley using (manipulated) natural voices, synthetic voices, and pathological voices. In addition, it is

investigated whether the uncanniness of voices can be explained by deviation from familiar categories or categorical ambiguity

Research question and hypotheses

First, the role of familiarity is investigated by comparing the effects of distortion on uncanniness for very familiar (human) voices and less familiar (cat) voices. Hypothesis 1 is thus:

1. Distortion of human voices increases uncanniness more than distortion of cat voices.

Furthermore, a vocal uncanny valley is replicated, including artificially distorted and naturally pathological voice stimuli. It is tested whether a vocal uncanny valley exists in principle but is successfully avoided by contemporary synthetic voices. Hypotheses 2 and 3 are thus:

2. A monotonic function of human likeness can best explain the uncanniness of synthetic and natural voices.
3. A non-monotonic function akin to an uncanny valley can best explain the uncanniness of synthetic, natural, distorted, and pathological voices.

Finally, it is investigated whether ambiguity in categorizing a voice as either human or non-human can best explain the uncanniness ratings. Categorization ambiguity is operationalized as 1. Categorization reaction time, and 2. Categorization uncertainty, i.e., the inconsistency of categorizations across participants. Hypothesis 4 is thus:

4. Categorization reaction time and categorization uncertainty predict uncanniness ratings of voices.

Methods

Participants. Power analysis revealed that $n = 50$ participants are sufficient to exceed a power of $1–\beta = 0.8$ with a six-voice-conditions within-subject design and a standard effect size of $d = 0.5$ (Cohen, 1988). Participants were Psychology students at the Cardiff University School of Psychology, recruited via the Experimental Management System (EMS). Participants were on average 19 years old ($SD_{age} = 1.05$), 37 identified as female, 11 as male, one as other, and one preferred not to say. Participants were compensated with 4 credits equivalent to the advertised compensation of a 60 minutes online study.

Stimuli. Ten typical and 15 pathological voices were taken from the Perceptual Voice Qualities Database (PVQD; Walden, 2022). Specifically, the 15 pathological voices with the highest subjective severity ratings were selected as stimuli. Specific pathologies included Reinke's Edema (x3), lesions (x3), vocal fold paralysis (x3), muscle tension dysphonia (x2), ulcerative laryngitis, adductor spasmodic dysphoria, and one unrecorded pathology. Ten distorted voice variants were created by using the STRAIGHT software, specifically by multiplying the normal voices' fundamental frequencies by 1000 (Kawahara et al., 2008). In addition, 10 normal cat meowing sounds were selected from www.freesound.org, and 10 distorted variants were created with STRAIGHT by multiplying the fundamental frequency by 1000. Finally, 15 synthetic voices were selected from various sources: Four mechanical sounds were taken from www.freesound.org, five voices from IBM Watson, three voices by Azure Microsoft TTS, Microsoft Sam, one voice created by a Stephen Hawking Voice Generator, and one generic Google TTS voice.

Fifteen pathological and synthetic voices were selected instead of 10 (as in the other conditions) because both conditions were expected to be more heterogenous and thus would need a higher stimulus number to be adequately statistically represented.

Because the fifteen pathological voices consisted of 9 female and 6 male voices, the same ratio was selected for typical voice counterparts. As distorted voices were created by manipulating the typical voices, the same ratio was present for those. For synthetic voices, six were artificial female voices, five were artificial male voices, and four were mechanical sounds. Voice accent was not controlled.

All stimuli were shortened to be around 5 seconds in length. For standardization, all typical, pathological, distorted, and synthetic voices (except for the mechanical sounds) were expressing the same sentences. The spoken sentences, “The blue spot is on the key again. How hard did he hit him?”, were used as basic sentences in the PVQD database and recreated for synthetic voices. More details on the voice stimuli are shown in *Table A4*.

Rating Task. For the rating task, participants had to rate each sound based on three items: eerie/uncanny, strange/weird, and humanlike/realistic. Items ranged from the extremes of 0 to 100 and participants could choose to place the slider on any point of the item. Voices were presented in a random order for each participant and were replayed for each item. Participants had an unlimited amount of time responding to the items. Because uncanniness and human likeness are here understood as subjective experiences and assessments, the terms were presented with minimal information to the participants to gauge their own interpretations.

The eerie/uncanny and strange/weird items were combined into an uncanniness index by calculating the means of the items for each stimulus. Analogous item combinations have been used in previous research with reliable consistency (Diel et al., 2022; Ho & MacDorman, 2017; Kätsyri et al., 2017).

Categorization Task. For the categorization task, all sounds except the normal and distorted cat meow voices were used. For each presented sound, participants had to do a two-alternative forced choice task on whether the voice was humanlike or not. Participants first

heard 2 seconds of the sound before the choice text appeared, at which point participants had the ability to decide by pressing either the left or right key on their keyboard. Participants were instructed to be as accurate and fast as possible.

Procedure. The whole procedure was conducted online. After giving informed consent and filling out a demographic questionnaire asking for participants' gender and age, participants were redirected to the experiment. They first went through the rating task followed by the categorization task.

The human likeness ratings were used to operationalize the x-axis of the uncanny valley function. Meanwhile, human categorization responses (both reaction times and response inconsistencies) were used as indicators of categorical ambiguity.

Analysis, ethics statement, and data availability. Analysis was conducted via R. Linear mixed models were used to control for participants, as well as analyses of variance (ANOVAs) and linear regressions. Data cleaning was conducted by removing all outlier (1.5*IQR) uncanniness, human likeness, and categorization reaction time ratings for each stimulus. A total of 17 values were removed. The experiment was approved by the Cardiff University School of Psychology Ethics Committee in October 2021 (reference number: EC.21.09.14.6411G). All methods were performed in accordance with the Declaration of Helsinki and informed consent was collected from all participants. at the datasets generated and analysed during the current studies and the analysis scripts are available on OSF: https://osf.io/7xs6j. Original versions of the voice samples can be downloaded from the PVQD website.

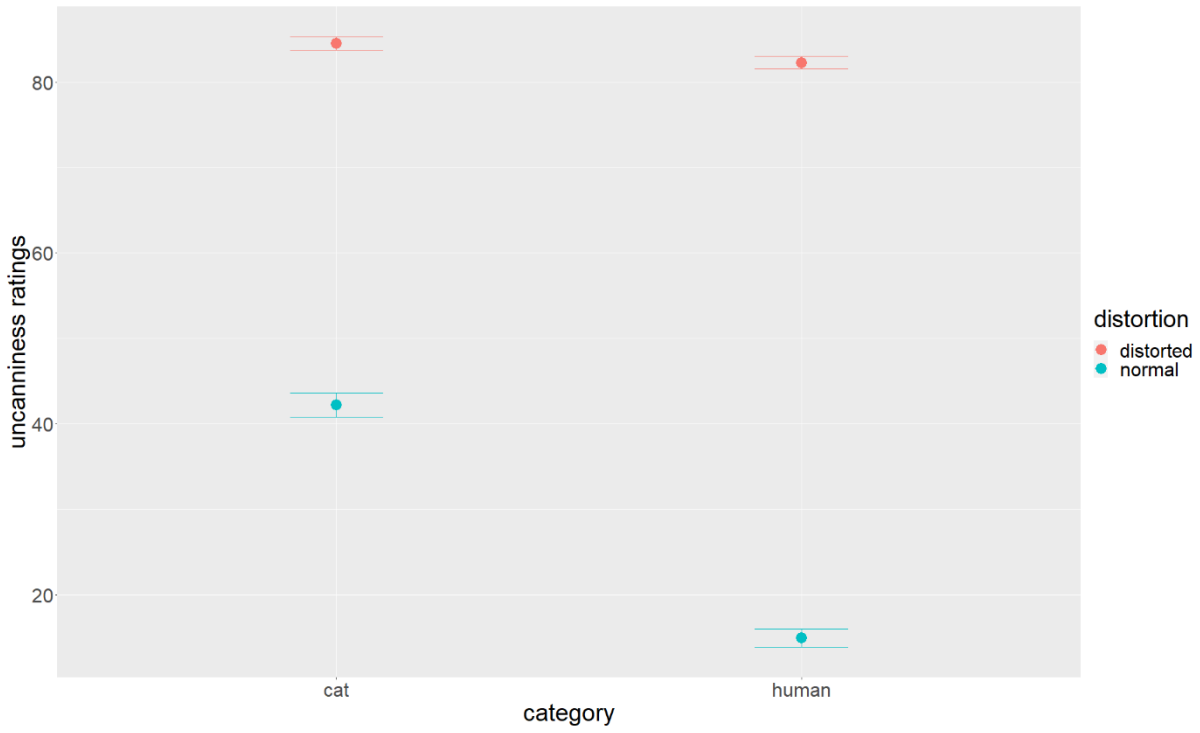
Results

Rating scales. The *eerie/uncanny* and *strange/weird* items were combined into an *uncanniness* index with a Cronbach's alpha of $\alpha = .79$, indicating acceptable, almost good construct validity.

Voice distortion: human vs cat. A within-subject 2 × 2 ANOVA was conducted with distortion (normal vs distorted) and species (cat vs human) as factors of uncanniness. The analysis showed main effects of both distortion ($F(1,48) = 567.02$, $p < .001$, $d = .77$) and species ($F(1,48) = 51.84$, $p < .001$, $d = .20$), as well as an interaction between these two ($F(1,48) = 47.35$, $p < .001$, $d = .15$). The interaction is visualized in *Figure 7.1*.

Figure 7.1

Average uncanniness ratings of distorted or undistorted cat and human voices. Error bars indicate standard errors.



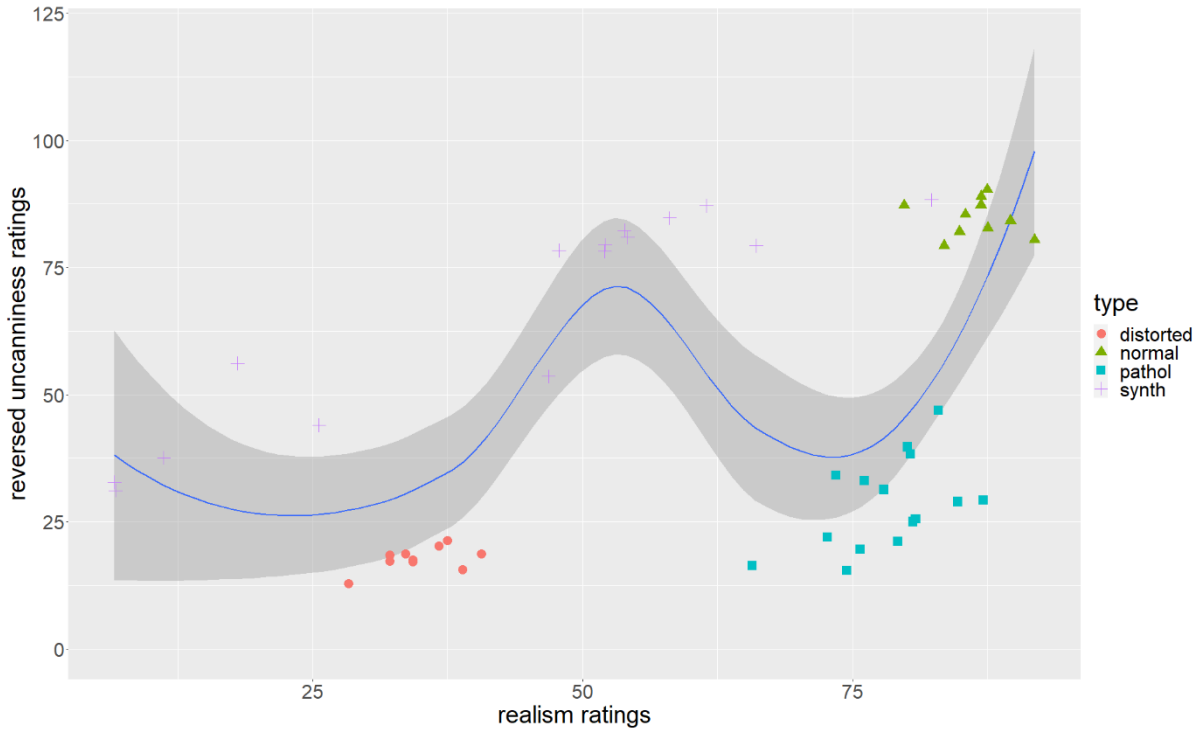
Follow-up *p*-adjusted post-hoc Tukey tests showed that distortion increased the uncanniness of both cat ($t(1825) = 33$, $p_{adj} < .001$, $d = 2.16$, CI[0.8, 3.52]) and human voices

($t(1825) = 52.48$, $p_{adj} < .001$, $d = 3.43$, CI[1.27, 5.59]). Furthermore, normal human voices were significant less uncanny than cat voices ($t(1825) = 21.328$, $p_{adj} < .001$, $d = 1.39$, CI[0.51, 2.27]), but not in the distortion conditions ($t(1825) = 1.82$, $p_{adj} = .19$, $d = 0.12$, CI[-0.03, 0.27]). Thus, the same distortion procedure increased the uncanniness of human voices more than the uncanniness of cat voices. Hypothesis 1 is thus supported.

An uncanny valley of voices. An uncanny valley of voice stimuli was investigated using a linear mixed model with human likeness ratings as fixed effects and participants and stimuli as random effects on uncanniness. Cat sounds were excluded from the analysis to focus on humanlike and mechanical voices. Results show that a cubic term ($t(1637) = -5.51$, $p < .001$, $R^2_{adj} = .67$) could explain the variance better than a linear term ($\chi^2 = 57.57$, $p < .001$) or a quadratic term ($\chi^2 = 30.27$, $p < .001$). The model is plotted in *Figure 7.2*.

Figure 7.2

Reversed uncanniness ratings plotted against realism ratings across voice type conditions. Dots indicate separate voice stimuli. The blue line shows a regression curve, and grey areas indicate standard errors.

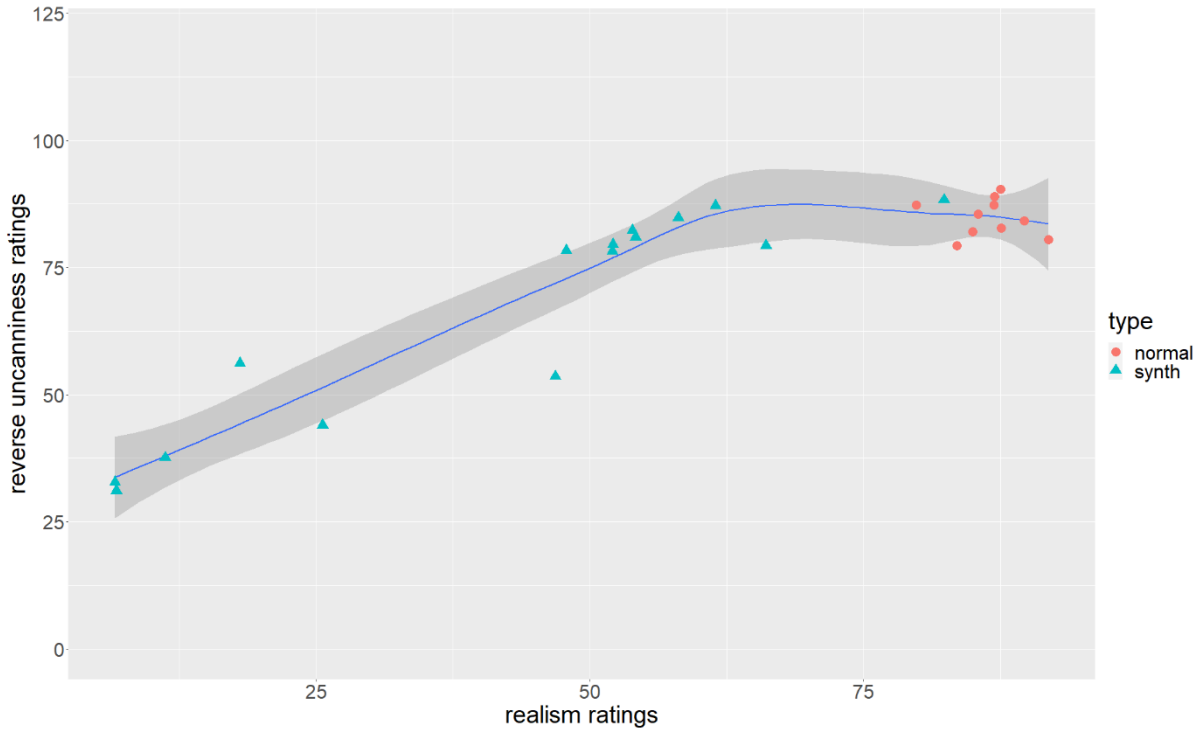


As can be seen in the plot, confidence intervals in the curves' "valleys" do not overlap with the confidence intervals of the curves' maxima. Taken together with the significant cubic term, a non-monotonic relationship explains the relationship between uncanniness and human likeness across voice categories.

In a second step, distorted and pathological voices were removed and the analysis was redone. The results show that again, a cubic term ($t(26000) = -2.86$, $p = .004$, $R^2_{adj} = .56$) could better explain the variance than a linear ($\chi^2 = 47.62$, $p < .001$) or quadratic term ($\chi^2 = 8.16$, $p = .004$). The function, depicted in *Figure 7.3*, however does not reflect an uncanny valley plot.

Figure 7.3

Reversed uncanniness ratings plotted against realism ratings across voice type conditions but excluding distorted and pathological voices. Dots indicate separate voice stimuli. The blue line shows a regression curve, and grey areas indicate standard errors.

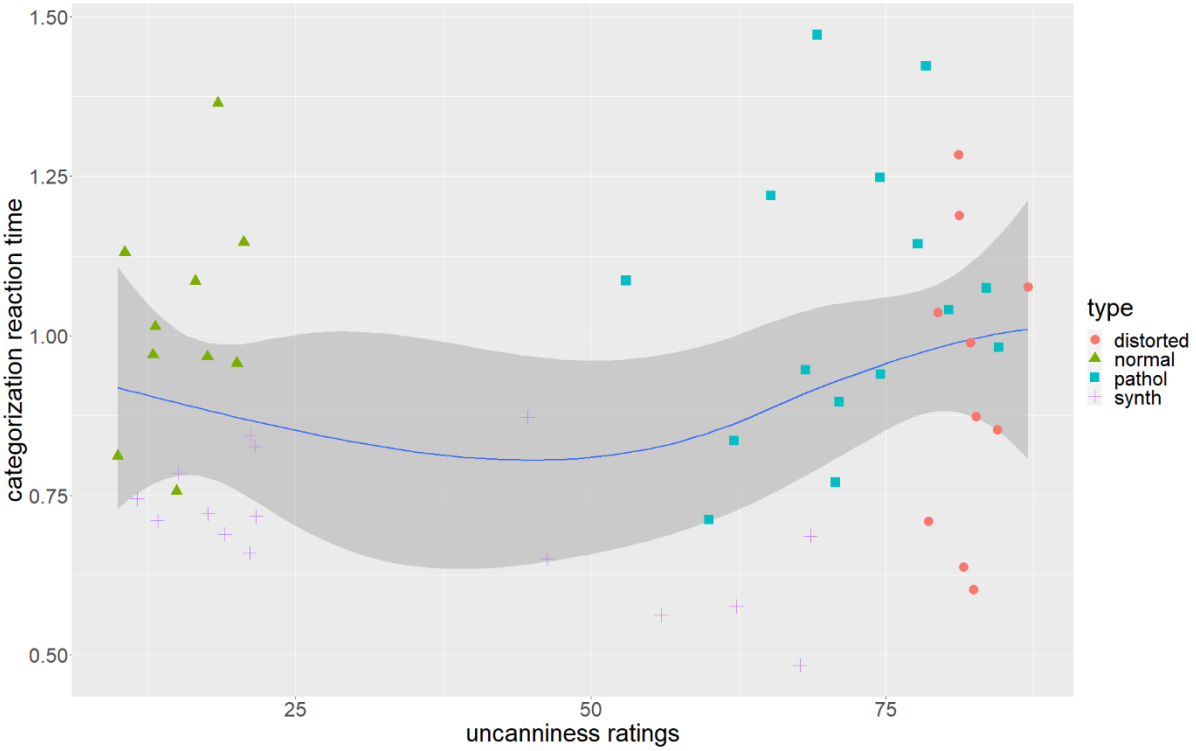


To complement the function, a second plot with distorted, normal, and pathological voices was plotted as well. Given that at no point in the functions in *Figure 7.3*, the confidence intervals seem to significantly decrease, but only increase with increasing realism, both functions indicate monotonic relationships between uncanniness and human likeness when the data depicted in *Figure 7.2* is divided based on different voice categories. Thus, a non-monotonic relation between uncanniness and human likeness seems to result from a combination of multiple monotonic functions. Thus, hypotheses 2 and 3 are supported.

Categorization difficulty as a predictor of voice uncanniness. A linear mixed model with reaction time as a fixed effect and stimuli and participants as random effects on uncanniness showed that reaction time could not predict voice uncanniness ratings ($t(2207) = 1.29$, $p = .197$). The data is plotted in *Figure 7.4*.

Figure 7.4

Categorization reaction time plotted against uncanniness ratings across voice type conditions. Dots indicate separate voice stimuli. The blue line shows a regression curve, and grey areas indicate standard errors.

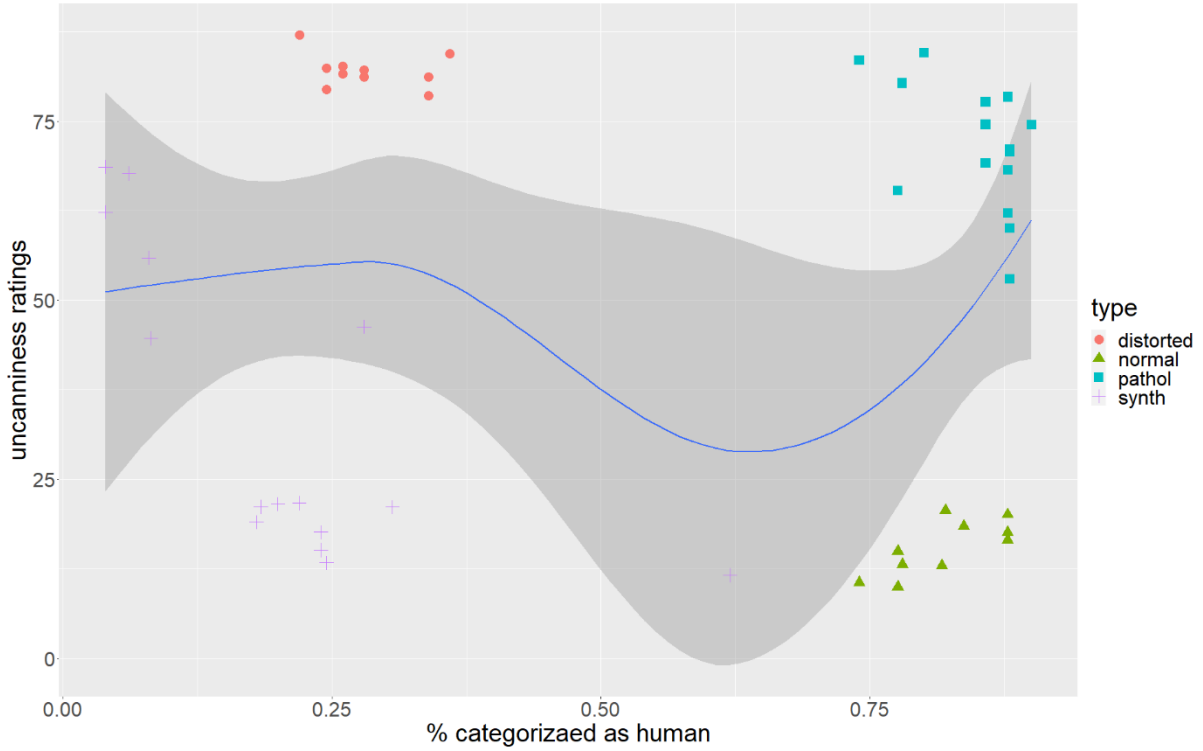


Voice categorization data was transformed into a *voice certainty* variable by coding participants' *non-human* categorizations as 0 and *human* categorizations as 1, then using the absolute values after subtracting the averaged categorizations for each stimulus by 0.5. *Voice certainty* thus reflects a variable ranging from 0 (50:50 categorization as human and non-human across participants) to 0.5 (consistent categorization as either human or non-human across participants) to be used as an operationalization of consistent categorization.

Because the transformed data was already aggregated across participants for each stimulus, a linear regression model was used to investigate the effect of categorization certainty on uncanniness. The results show that categorization certainty could not predict voice uncanniness ratings ($t(50) = 0.15$, $p = .88$), and the data is visualized in *Figure 7.5*.

Figure 7.5

Reversed uncanniness ratings plotted against across-participant percentage at which stimuli were categorized as human, across voice type conditions. Dots indicate separate voice stimuli. The blue line shows a regression curve, and grey areas indicate standard errors.



The figures indicate that while distorted voices were both ambiguous and uncanny (compared to synthetic and normal voices which were neither ambiguous nor uncanny), pathological voices seemed to be uncanny yet consistently categorized as human (Fig. 6). This has been mostly confirmed by post-hoc tests: While distorted voices were not more ambiguous and uncanny than synthetic (ambiguous: $t(46) = -4.553$, $p < .001$; uncanny: $t(46) = 9.192$, $p < .001$) or human voices (ambiguous: $t(46) = -3.197$, $p = .008$; uncanny: $t(46) = 11.59$, $p < .001$), pathological voices were more uncanny than synthetic ($t(46) = 8.03$, $p < .001$) and human voices ($t(46) = 10.69$, $p < .001$), while not being more ambiguous (synthetic: $t(46) = 1.29$, $p = .475$; human: $t(46) = -1.05$, $p = .621$). Thus, hypothesis 4 was not supported.

Human categorization as a moderator of human-deviation on uncanniness. The model plotted in Fig. 2 indicates a *W*-shaped relationship with “two valleys”. Such a relationship may be a consequence of choosing different categories of voices which interact differently with human likeness to affect uncanniness. The effect of voice type could be moderated by a variable influencing the perception of a decrease in realism (or closeness to the human norm) on uncanniness. Hence, a third variable may underlie the observed data by moderating a linear relationship between human likeness and uncanniness. As the uncanny valley has been linked to perceptions of markers of death and disease avoidance (MacDorman & Ishiguro, 2006; Mori, 2012), the effect may be linked to the perception of organic appearance. ,. Thus, a perceived high “organicness” of a voice may increase the sensitivity of uncanniness towards deviations from human likeness, potentially due to evolutionary disease avoidance mechanism. Although “organicness” has not been measured in the experiment, the categorization of a voice as human may indicate how organic it was perceived to be, as both distorted and synthetic voices were categorized as non-human, while pathological and normal voices were categorized as human. Categorization of a stimulus as human may increase the effect of deviation on uncanniness: Hence, the slope from fully synthetic to human voices would be less steep than for (partially artificial) distorted to human voices, which would be again less steep than the slope for (fully organic) pathological to human voices. A post-hoc linear regression analysis has thus been conducted for the interaction between categorization response (human vs non-human) and human likeness on uncanniness. The results show main effects of response ($t(46) = 10.011$, $p < .001$), human likeness (($t(46) = -8.922$, $p < .001$), and an interaction between these two (($t(46) = -6.163$, $p < .001$; $R^2_{adj} = .80$). Thus, a moderated linear relationship between human likeness, uncanniness, and categorization as “human” is indicated.

Discussion

Voice distortion and familiarity. Voice distortion created by multiplying the fundamental frequency by 1000 increased the uncanniness of both human cat voices. The increase was stronger for human compared to cat voices. A higher degree of familiarity to a voice category may sensitize uncanniness caused by deviation.

Differences in fine details between human voices carry vital information about spoken messages and characteristics and states of the speaker (Kreiman et al., 2018; Kreiman & Gerratt, 1992). The recognition of analogous information is less important for the perception of cat voices. Thus, the degree of familiarity (and change sensitivity) in humans is lower for cat compared to human sounds. Higher uncanniness sensitivity for human compared to cat voices can thus be explained by higher familiarity to typical voice patterns and sensitivity to deviations from these patterns.

An uncanny valley of voices. A function with only synthetic and normal human voices showed that a linear relationship between human likeness and uncanniness akin to previous research (Baird et al., 2018a, 2018b; Kimura & Yotsumoto, 2018; Kühne et al., 2020; Romportl, 2014). However, adding voices that are either deliberately distorted or naturally deviating produces a non-monotonic function of uncanniness and human likeness. Especially when excluding distorted voices, the curve would be akin to an *N*-shaped uncanny valley plot, and the pathological voices would lie within an uncanny valley akin to the prediction of dead bodies falling into an uncanny valley (Mori, 2012). Such an interpretation would favour explanations of the uncanny valley related to mortality salience or disease avoidance (MacDorman & 2006).

Previous researchers have noted that an uncanny valley could occur at any point at a graph, allowing multiple valley-shaped functions, potentially due to a multicausal emergence of the

effect (Diel & MacDorman, 2021; Bartneck et al., 2009; Hanson, 2006; Kim et al., 2022). An uncanny valley may not necessarily occur on just one area on the human likeness axis, and polynomial functions more complex than an N-shaped curve may occur depending on the stimuli selected, as in this study.

Categorization ambiguity does not predict uncanniness. Categorization ambiguity has been proposed to underlie the uncanny valley effect (Cheetham et al., 2013; Weis & Wiese, 2017; Yamada et al., 2013). This study failed to find evidence for the categorization ambiguity hypothesis: Neither categorization reaction time nor categorization response consistency could predict uncanniness ratings. While distorted voices were both uncanny and difficult to categorize, pathological voices were not. Categorization ambiguity may correlate with stimulus deviations when stimuli are incremental morphs between two easily categorizable stimuli (Yamada et al., 2013) and thus may be uncanny due to their deviation. However, certain stimuli can be uncanny despite being easy to categorize (Mathur et al., 2020; Diel & MacDorman, 2021). Thus, uncanniness cannot be explained solely by categorization ambiguity.

A moderator of uncanniness? A significant interaction between human categorization and human likeness was found that could explain uncanniness better than a polynomial model of human likeness. Categorization as human sensitized the effect of deviation on uncanniness. As, dehumanization can decrease the uncanniness of androids (Yam, Bigman, & Gray, 2021), categorization as human may activate a stricter evaluation of stimuli based on their proximity to the human norm. As a humanization manipulation can affect the specialized processing of faces (Fincher & Tetlock, 2016; Fincher, Tetlock, & Morris, 2017), an increase of humanization (and human categorization) may also further sensitize the detection of configural deviations and thus uncanniness. Similarly, as mind perception increases configural processing (Deska, Almaraz, & Hugenberg, 2017), it may also increase the

sensitivity to deviations and thus uncanniness when a stimulus is perceived as both having a mind and deviates from the norm of appearance.

However, human likeness and categorization choice was highly correlated in this study, decoupling human likeness (or deviation) from human categorization (or humanization) would be required, which however should be difficult given the conceptual similarity of these concepts.

Experiment 11

Research Question and hypotheses

The aim of Experiment 11 is to investigate a potential third variable that may moderate a monotonic effect of human likeness on uncanniness. Several candidates for this third variable were explored.

Pathogen avoidance: Perception of organic voice. Uncanniness may be a response to the detection of indicators of contagious disease (MacDorman & Entezari, 2015; MacDorman & Ishiguro, 2006). Disease indicators may appear as physical anomalies or deviations co-occurring with pathology or physical disabilities (Schaller, Paul & Faulkner, 2003; Workman et al., 2021). As disease threat is only relevant for organic material, the perception of an entity being organic (vs synthetic) should then increase negative response towards norm deviation in a stimulus. Meanwhile, a voice recognized as inorganic should pose no disease-related threat even despite deviating from the norm.

1. Perception of organicness moderates the relation between human likeness and uncanniness across voice categories

Mind attribution and animacy. Uncanniness may be elicited when human qualities like mind or animacy are attributed to non-human entities (Gray & Wegner, 2012; Stein & Ohler, 2017). Thus, less humanlike voices not perceived as having a mind or being animate should

not elicit uncanniness, while deviating voices which appear to have a mind or to be animate should be uncanny.

2. Attribution of mind moderates the relation between human likeness and uncanniness across voice categories
3. Perception of animacy moderates the relation between human likeness and uncanniness across voice categories

Method

Participants. According to a power analysis, $n = 35$ participants are sufficient to exceed a power of $1\text{-}\beta = 0.8$ for a within-subject design with a standard effect size of $d = 0.5$ (Cohen, 1988). Participants were Psychology students at the Cardiff University School of Psychology, recruited via the Experimental Management System (EMS). Participants were on average 19.26 years old ($SD_{age} = 1.29$), 34 identified as female and one as male.

Stimuli. Per category (distorted, normal, pathological, synthetic), five stimuli were selected from Experiment 1. In addition, variation of distortion degree was created for distorted and pathological voices: For distorted voices, fundamental frequencies of normal (base) voices were increased by 250, 500 and 750, in addition to the present distorted voices with an increase by the value of 1000. These distortion levels were created to simulate an incremental increase of distortions starting with the normal counterparts. As the goal of the experiment is to investigate a moderated linear function of uncanniness, an incremental increase of distortion may reflect a linear function for one value of the moderator variable. For pathological voices, additional sets of five voices were selected based on the level of perceived severity ratings as reported in the PVQD (Walden, 2022). The five most severe pathological voices were selected for the severity rating limits of 100, 75, 50, and 25. Spoken sentences were the same as in Experiment 1. The stimuli are summarized in *Table A5*.

Procedure: Rating Task. The experiment consisted only of a rating task conducted online. The rating task was identical to the one in Experiment 10, except participants rated each voice based on the items eerie, strange, and humanlike only, in addition to its perceived animacy, mind attribution, and organicness. The additional items were presented the same way as the previous ones described in Experiment 10.

Analysis, ethics statement, and data availability. Analysis was conducted via R. Linear mixed models were used to control for participants, as well as analyses of variance (ANOVAs) and linear regressions. Data cleaning was conducted by removing all outlier (1.5*IQR) uncanniness, human likeness, and categorization reaction time ratings for each stimulus. A total of 13 values were removed. All methods were performed in accordance with the Declaration of Helsinki and informed consent was collected from all participants. at the datasets generated and analysed during the current studies and the analysis scripts are available on OSF: https://osf.io/7xs6j.

Results

Rating scales. Eerie and *strange* items were combined into an *uncanniness* index with a Cronbach's alpha of $\alpha = .8$, indicating good consistency.

Moderating effects. Linear mixed models with human likeness and either animacy, mind attribution, or organicness as fixed effects stimuli and participants as random effects showed that the interaction between human likeness and animacy ($t(1762) = -3.568$, $p < .001$, $R^2_{adj} = .58$), mind attribution ($t(1856) = 2.824$, $p = -005$, $R^2_{adj} = .57$), or organicness ($t(1690) = -2.539$, $p = .011$, $R^2_{adj} = .58$) each significantly predicted uncanniness.

To test whether a moderated function can explain uncanniness better than a quadratic function of human likeness, the linear moderator models were tested against a quadratic human likeness function. A quadratic human likeness model was able to predict uncanniness

($t(6366) = -4.065$, $p < .001$, $R^2_{adj} = .56$). Model comparisons showed that only the model with organicness fitted the data significantly better than the quadratic human likeness model ($\chi^2 = 20.184$, $p < .001$). Replacing a quadratic human likeness term with either animacy or mind perception did not change model fit. Thus, a moderated linear function of organicness and human likeness could explain the results better than a quadratic function of human likeness.

Differences between voice types. P-adjusted Tukey tests on differences between voice categories showed that distorted voices were more uncanny than normal ($t(56) = 6.789$, $p_{adj} < .001$) and synthetic voices ($t(56) = 7.097$, $p_{adj} < .001$). However, while distorted voices were perceived as less animate ($t(55) = -9.825$, $p_{adj} < .001$) and as having less mind ($t(55) = -9.725$, $p_{adj} < .001$) compared to normal voices, they did not differ from synthetic voices.

Discussion

"Uncanny valley" as a moderated linear function. A third variable of organicness moderates a linear relationship between human likeness and uncanniness. A moderating function may appear as an increase of the slope with increasing organicness: While distinctively artificial voices can deviate from the human norm without suffering from uncanniness, deviations in organic-sounding voices may quickly become unnerving, for example due to the threat of contamination from infected organic entities (MacDorman & Ishiguro, 2006).

However, all tested predictors were highly intercorrelated, and correlated highly with human likeness. Thus, it is not clear whether organicness itself is the third variable, or whether the third variable can be better described by a different construct.

Animacy and mind perception. Previous research aimed to explain the uncanny valley phenomenon through the attribution of humanlike characteristics like animacy or mind onto

visibly artificial or inanimate stimuli (e.g., Stein & Ohler, 2017). However, the present results suggest that voice uncanniness also occurs for deviating voices clearly perceived as animate or having a mind (i.e., pathological voices). Meanwhile, artificially distorted voices perceived as inanimate or lacking mind were still uncanny. These results cannot be explained by misattribution of human qualities onto artificial entities.

General Discussion

Uncanny valley of voices

In two experiments, non-monotonic relationships between uncanniness and human likeness for voices were observed, although the function differs from a typical uncanny valley function (Mori, 2012). The cognitive processing underlying the uncanny valley effect may be analogous across visual and auditory domains. Distinct face and voice variants elicit stronger activity in neural substrates specific to these categories (Andics et al., 2010; Latinus et al., 2010; Loffler et al., 2005), which may indicate increased processing need. Increased processing need may in turn decrease the aesthetic appeal of a stimulus (Winkielman et al., 2003). Alternatively, a higher familiarity with a face or voice category may sensitize to errors or deviations, leading to prediction error signals (Friston & Kiebel, 2011; Saygin et al., 2012).

Synthetic voices and the uncanny valley

Synthetic voices did not fall into the valley of the function and instead were allocated around it Hence, modern TTS synthetisation can successfully replicate human voices. In fact, participants consistently rated one of the Watson voices to be about as humanlike as typical human voices (however, the same voice was ambiguously categorized with a 53% human categorization rate). Thus, synthetic voices manage to overcome the uncanny valley while visual synthetic replications of humans (i.e., androids) often do not.

It may be easier to replicate a synthetic voice than a synthetic face without errors: Synthetic voice replication can rely on recorded natural voices while synthetic faces must be artificially reconstructed. Alternatively, as human identity discrimination ability is more sensitive to faces than to voices (Barsics, 2014), visual human processing may also be more sensitive to deviations compared to auditory human processing, making errors in design more apparent and appalling.

In general, the results affirm current technology of artificial voice: While a vocal uncanny valley exist, today's artificial voices manage to overcome it.

Theories on the uncanny valley

The present results conflict with two existing theories on the uncanny valley: That uncanniness is caused by either 1) categorical ambiguity or categorization difficulty, or 2) by misattribution of human qualities onto nonhuman entities. While distorted voices in Experiment 10 were both uncanny and categorically ambiguous, pathological voices were uncanny despite being clearly categorized as human. In Experiment 11, distorted voices were uncanny despite having less mind or animacy attributed to them than normal voices, and with no differences compared to synthetic voices. Furthermore, pathological voices were uncanny in both experiments, contrasting the misattribution theory's prediction that uncanniness is caused by non-human entities.

The present data can be better explained by a deviation-from-familiarity account (Chapters 2-4): both distorted and pathological voices are uncanny because they deviate from the pattern of human voice that has been experienced throughout life. Categorical ambiguity can correlate with stimulus uncanniness as categorically ambiguous stimuli (Yamada et al., 2013) also deviate from typical appearance. Similarly, mind attribution can enhance configural processing of faces (Deska et al., 2017), which in turn may sensitize the negative evaluation

of deviations (Chapters 2 to 4). Thus, mind attribution may increase uncanniness by sensitizing to deviations (Müller et al., 2020; Yam, Bigman, & Gray, 2021; Yin, Wang, Guo, & Shao, 2021). The interaction between attribution of human qualities, degree of configural processing, and uncanniness sensitivity can be explored in future research.

A moderated monotonic function of uncanniness

Rather than being a non-monotonic, valley-shaped function, the uncanny valley may consist of two or more monotonic functions with different slopes (e.g., one for an increase of likability from synthetic to full human variants, and one for a decrease of uncanniness from deviating or abnormal to typical humanlike variants). To test this, both experiments have investigated a moderated linear function of uncanniness.

Experiment 10 found that a moderated linear function could predict uncanniness, and Experiment 11 found that it could explain uncanniness better than a non-linear function of human likeness. Although the specific moderating variables differed between experiments, both "human" categorization and perceived organicness increased the effect of deviation on uncanniness. However, both variables also highly correlated with human likeness.

The investigated moderator variables are evolutionarily sensible: Disease avoidance may underlie the uncanny valley effect (MacDorman & Entezari, 2015), and markers of infectious disease are expressed as changes from typical (human) appearance or behaviour (Schaller et al., 2003). Given that the threat of infection is present only in organic entities, avoidance of deviating organic or human entities should be effective for minimizing risk of infection. Meanwhile deviating yet clearly inorganic entities pose no threat of infection.

Alternatively, the increased uncanniness for less humanlike stimuli in organic entities or those categorized as hu man may be due to a higher level of perceptual experience with

naturally humanlike stimuli: Perceptual expertise with a stimulus category increases the uncanniness of deviating exemplars (Chapters 2 to 4).

Limitations and future directions

Interpretations of test results on a moderated linear function of the uncanny valley are limited due to the intercorrelation between the predictors. As multicollinearity cannot be excluded, the exact relationship between the predictor variables and uncanniness remains unclear. Future research may aim to tackle this problem using decorrelated predictors.

The use of linguistic content in the stimuli adds additional dimensions which could have influenced the results. For the difference between distortion effects on human and cat voices, a reduced intelligibility of the human voices but not cat voices due to distortion may have been a reason for the increased uncanniness for distorted human voices. Similarly, as distorted and pathological voices could be less intelligible, the additional processing need for these voices could have been a cause of uncanniness.

Chapter 7 presented research testing predictions of categorization ambiguity, mind/animacy attribution, and the refined theory (linked to disease avoidance) against each other. While uncanny stimuli were not necessarily ambiguous or had human qualities attributed onto them (nor were ambiguous stimuli or those with human qualities attributed to them necessarily uncanny), deviating stimuli tended to be uncanny, especially in familiar (human vs cat stimuli) and biological (organic vs robotic) stimuli. Hence, Evidence was found only for the latter explanation. Chapter 8 presents further research critically investigating theories of the uncanny valley, specifically by using an affective priming paradigm to investigate disease avoidance and mortality salience hypotheses. Furthermore, an inversion paradigm is used to investigate the refined theory in further categories, namely motion and body stimuli.

Chapter 8: Evidence against disease avoidance and mortality salience explanations of the uncanny valley, partial evidence for configural processing

Methods, experiments, and large portions of the introduction and discussion in this chapter are currently in review in the journal *Cognition*.

Introduction

The present chapter presents an emotional priming experiment to investigate disease-avoidance and mortality salience theories of the uncanny valley, two theories that have received little to no critical attention. In addition, the refined theory is critically investigated by testing for inversion effects in videos of ecologically relevant uncanny android and CG characters.

Deviation from specialized categories

The effect of deviation in specialized categories has not yet been investigated in body stimuli, nor in stimuli depicting biological motion. However, the *Thatcher illusion* has been observed in motion (Mirenzi & Hiris, 2011; Schwaninger & Cunningham, 2002), and an inversion effect has been found for bodies (Keye, Mingming, Tiantian, Wenbo, & Weiqi, 2019; Reed, Stone, Bozova, & Tanaka, 2003; Stekelenburg & de Gelder, 2004). Thus, a reduction of uncanniness caused by configural deviation through inversion is expected in these categories as well.

Disease and threat avoidance

Mori (2012) first suggested that the uncanniness of humanlike entities may stem from their similarities to dead or diseased human bodies. The *disease avoidance* theory explains the uncanny valley as a reaction towards indicators of disease (MacDorman & Ishiguro, 2006). Disgust has been associated with the uncanniness of humanlike entities in past research (Ho, MacDorman, & Pramono, 2008; MacDorman & Entezari, 2015). As disgust can be

understood as an evolutionary beneficial response towards the avoidance of contamination (Rozin & Fallon, 1987), the association between uncanny stimuli and disgust supports the idea of evolutionary disease avoidance mechanisms. *Danger avoidance* (Moosa & Ud-Dean, 2010) further extends the explanation to include negative reactions towards dead bodies as threats beyond contamination (e.g., whatever was responsible for the death of the observed organism). Both explanations rely on the uncanny stimulus being associated with death or disease—at least unconsciously. Thus, automatic reactions towards uncanny stimuli should be comparable with those towards disgust-eliciting stimuli.

Mortality salience

It has been suggested that a dislike of artificial humanlike entities is caused by reminders of one's own mortality (*mortality salience*; MacDorman, 2005; Koschate, Potter, Bremner, & Levine, 2016). *Terror management theory* predicts that mortality salience leads to the activation unconscious defence mechanisms to reduce anxiety and promote self-preservation (Greenberg, Pyszynski, & Solomon, 1986; Pyszczynski, Solomon, & Greenberg, 2015). MacDorman (2005) found that viewing uncanny robots increased preference for people who support one's worldview, as predicted by terror management theory (Greenberg, Pyszynski, & Solomon, 1986). In addition, Koschate et al. (2016) found that uncanny stimuli increased the accessibility to death-related thoughts. Thus, the available yet sparse research supports the mortality salience explanation of the uncanny valley.

However, the relation between uncanniness and mortality salience remains unclear: Previous research generally manipulated mortality salience by using death-related or control stimuli to investigate their effect on, for example, death-thought accessibility (Pyszczynski, Solomon, & Greenberg, 2015). However, a measured difference on a mortality salience dependent variable does not necessarily indicate mortality salience because the difference could have been caused by other factors. For example, uncanny androids could activate death-related

concepts not because they resemble dead people but because they elicit anxiety-related reactions that in turn activate death-related thoughts. Hence, further research is needed to investigate the association between uncanny stimuli and mortality salience.

Mortality priming activates associations of death, increasing accessibility of death-related concepts, observable as faster reaction times for death-related words than for control words (Hayes, Schimel, Arndt, & Faucher, 2010; Huang & Wyer, 2015). Thus, uncanny stimuli should also activate death-related concepts if the mortality salience explanation of the uncanny valley were correct.

One way to investigate the effect of an emotional reaction on the activation of semantic concepts is through priming: presenting emotion-inducing pictures and measuring reaction towards stimuli representing relevant concepts (Neuman & Lozo, 2012). Reaction times towards target words in a lexical decision task is often used to investigate priming effects; however, the direction of change (increased or decreased reaction time) can differ between studies: It is often presumed that a prime enhances the ability to activate semantically associated activation, leading to faster reaction times (Fazio, Sanbonmatsu, Powell, & Kardes, 1986). Concordantly, Neuman and Lozo (2012) found that disgust-related stimuli decrease reaction time needed to correctly categorize a congruent emotional picture, indicating that emotional priming may activate concepts related to the specific emotion. Meanwhile, research on valence-dependent reaction times is mixed: While some studies show increased reaction times for negative compared to neutral words (e.g., Algom, Chajut, & Lev, 2004), others find deacreased reaction times (e.g., Kousta, Vinson, & Vigliocco, 2009; Scott, O'Donnell, & Sereno, 2014; Yap & Seow, 2014). Similar discrepancies are observed in emotional priming paradigms (e.g., Challis & Krane, 1988; Pan et al., 2016; Schmitz & Wentura, 2012; Spruyt, De Houwer, Hermans, & Eelen, 2007; Topolinski & Deutsch, 2013; Yao & Wang, 2013; Yao, Zhu, & Luo, 2019). For example, Yao et al. (2019) found that

reaction times were lower for congruent compared to incongruent trials following positive-valence primes while reaction times were increased in congruent compared to incongruent trials for negative-valence primes. An increase of reaction time for emotionally congruent stimuli after a negative-valence prime may result from an inhibitory response or motor suppression following the negative emotional prime (Clore & Stobeck, 2006). Alternatively, Estes and Adelman (2008; see also Yao et al., 2019) proposed that automatic vigilance may increase reaction time of negative words which may hold attention for longer than positive or neutral words. While previous research thus shows effects of negative emotional target words, the direction of the effect is unclear. Finally, no study as of yet has investigated the effects of uncanny primes specifically; hence, directional predictions on reaction times are unclear. In any case, if the uncanny valley is related to disease avoidance, an uncanny prime should show reactions to disease-related worlds analogous to a disease or disgust prime. Similarly, if the uncanny valley is related to mortality salience, an uncanny prime should elicit reaction time changes comparable to a threat or death prime. In summary, while the direction of an uncanny prime on semantic target words is unclear, it should reflect the effect of other relevant emotional primes like disgust (if related to disease avoidance) or fear (if related to mortality salience).

Experiment 12

This research is divided into two Experiments. Experiment 12 contains three parts: An uncanny valley replication including an inversion effect (Part 1), a priming study (Part 2), and a study in body distortions and inversion (Part 3). Experiment 13 is a more extensive replication of the priming study conducted with a different set of test stimuli and participants.

Hypotheses

In three parts, Experiment 12 empirically investigates the theories on the uncanny valley discussed above. If deviations from a familiar category cause the uncanny valley effect,

inversion of uncanny stimuli (e.g., android videos) should flatten the valley. The decrease of uncanniness for androids would be analogous to the decrease of uncanniness of moving Thatcher faces after inversion.

In addition, it is investigated whether the effect of inversion on uncanniness ratings is also present in normal and distorted bodies: Specifically, it is investigated whether distorting human bodies increases uncanniness, and whether this effect is reduced by inverting the stimulus. Thus, four complementary hypotheses are tested for the configural processing account.

First, because specialized processing is presumed to sensitize uncanniness in distorted stimuli, it is suggested that a nonlinear (quadratic or cubic) function of uncanniness and human likeness emerges only for upright (not inverted) stimuli:

- *Configural processing hypothesis 1:* Stimuli varying across the human likeness scale produce a non-linear (quadratic or cubic) uncanny valley shaped uncanniness function when upright, but not when inverted.

Second, if specialized processing sensitizes uncanniness, its disruption caused by inversion should recuse the uncanniness of uncanny stimuli, such as humanlike androids or distorted (Thatcherized) faces

- *Configural processing hypothesis 2:* Androids or computer-generated stimuli and moving Thatcherized faces are perceived as more uncanny when upright than when inverted.

Third, akin to previous research (Chapters 3 to 4), increasing distortion in bodies should increase uncanniness. However, as this effect is presumed to be sensitized by specialized processing, this increase should be reduced for inverted stimuli:

- *Configural processing hypothesis 3:* Increasing distortion of human bodies increases uncanniness ratings, but the effect is reduced when the bodies are inverted.

Finally, as it is presumed that specialized processing primarily causes a sensitization to deviating or atypical information which is then rated as uncanny, an inversion effect akin to the previous hypothesis is also expected for participants' tendency to categorize the stimuli as atypical:

- *Configural processing hypothesis 4*: Distortions of human bodies are more likely to be categorized as atypical when the stimuli are presented with increasing distortions, but this effect is reduced the when the bodies are inverted.

Mortality salience or disease and threat avoidance theories predict that uncanny stimuli elicit death or disease-related associations. Thus, if mechanisms for disease or threat avoidance or mortality salience underlie the uncanny valley effect, priming with uncanny stimuli should activate disease-related or death-related concepts more than non-uncanny stimuli. Finally, the effects of disease and fear primes should elicit similar effects on reaction times as uncanny primes. As the direction of the reaction time effect on negative (e.g., uncanny) primes is unclear (see above), more conservative, two-directional hypotheses are formed.

- *Death priming hypothesis 1*: Uncanny primes change reaction times for death-related words in relation to control words in a lexical decision task, compared to a neutral prime
    - *Death priming hypothesis 2*: Fear primes change reaction times for death-related words in relation to control words in a lexical decision task, compared to a neutral prime.

- *Disease priming hypothesis 1*: Uncanny primes change reaction times for disgust-related words in relation to control worlds in a lexical decision task, compared to a neutral prime
    - Disease priming hypothesis 2: Disgust primes change reaction times for disease-related words in relation to control words in a lexical decision task, compared to a neutral prime.

Methods

Experiment 12 is divided into three parts in which all participants took part in. We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

Participants. According to a power analysis using a medium effect size of $d = 0.5$ and using a 4 × 3 mixed design (four between-subject priming conditions, three within-subject word conditions), 132 participants (33 per priming condition) were sufficient for a power of $1 - \beta = .80$. Because no previous research using uncanny priming has been conducted, a standard medium effect size of $d = 0.5$ (Cohen, 1988) was used. Participants were undergraduate Psychology students, with an average age of $M_{age} = 19.73$ ($SD_{age} = 2.7$); 107 were female, 22 male, two other, and one participant preferred not to say.

The study was approved by the University's ethics committee (EC.21.09.14.6412G). This study was not preregistered.

Stimuli. For robot, android, or computer-generated (CG) stimuli, video clips from Ho and MacDorman (2017) were used because these stimuli have been previously validated as uncanny. Stimuli were a CG baby (Tin Toy, 1988) a CG man (Apology, 2008), a CG woman (Mary Smith from *Heavy Rain: The Casting*, 2006), a Roomba (iRobot), Kotaro (JSK, University of Tokyo), Jules (Hanson Robotics), and an android head (David Ng). In addition,

clips were added of the robots Pepper (SoftBank Robotics), Asimo (Honda), and Romeo (Aldebaran Robotics).

For Thatcherized faces, five video clips depicting frontal views of unemotional talking human faces from the FAMED database were used (Longmore & Tree, 2013). Thatcherized versions were created by inverting the eyes and mouth in the videos. All stimuli were presented either upright or inverted, creating a total of 20 stimuli (2 orientations × 2 conditions × 5 faces).

All clips were 15 seconds long and presented both upright and inverted. All stimuli were used in Study 1, and some of the stimuli were used in Part 2 of the experiment, for either the uncanny priming condition or the control condition. Specifically, the most uncanny stimuli were used for the priming condition, while only three out of five control (not Thatcherized) human videos were used because the stimuli were already very homogenous.

In the lexical decision task, participants decided whether word stimuli were real words. Stimuli were 48 letter strings out of which 8 were related to death or disgust or were neutral. The remaining 24 were nonwords. All words used are listed in *Table 2*. To investigate potential differences in word length of word frequency between the condition, two ANOVAs. However, no effects of conditions have been observed for either word length ($F(2,21) = 0.328$, $p = .724$) or frequency ($F(2,21) = 0.153$, $p = .859$). Thus, word length and frequency are comparable across conditions.

Table 8.1

Words used in the lexical decision task, divided by condition (word type).

Neutral words	Disease-related words	Death-related words	Nonwords

Book; essential; hallway; hound; mineral; quote; sandals; teacher	Contagious; infection; nausea; pest; rash; sick; ulcer; vomit	Coffin; deadly; doom; grave; killed; mortality; skull; tomb;	Actihro; afer; delliv; drivtt; falipi; glarst; gorpan; grusdi; holdok; horrk; kefft; kininal; krek; krin; midaun; mistisim; musear; ; perpe; roqua; sindoke; talal; tybs; uvalen; verrar

Five full body stimuli with censored faces were taken from the BEAST database (de Gelder & Van den Stock, 2011). The bodies were always standing upright, coloured in grey, and in various positions. Stimuli were incrementally distorted in five steps by elongating legs and shortening and displacing arms. Body stimuli were presented both upright and inverted, giving a total of 50 stimuli (5 bodies × 5 distortion levels × 2 orientations). All body stimuli were used in Part 3 of the experiment.

Finally, 10 fear-inducing and 10 disgust-inducing stimuli from the International Affective Picture System (IAPS, Lang, Bradley, & Cuthbert, 1999) were used as fear and disgust primes for Part 2 of the experiment.

Procedure. In the rating task, video clip stimuli were rated on the scales *weird*, *eerie*, and *humanlike*, each ranging from 0 to 100. Individual stimuli were presented simultaneously with each scale, and participants could take an unlimited amount of time for each scale. Each

video clip was played on repeat until rated. All stimuli were presented in random order. Scales were presented in a fixed order for each stimulus, and the scale *eerie* was reversed to reduce response bias (e.g., Weijters & Baumgartner, 2012).

After completing the rating task, participants were exposed to the priming manipulation, consisting of four conditions: uncanny, disgust, fear, and control. In the *uncanny* condition, participants viewed a 10s video consisting of short clips of the androids, CG characters, and Thatcher faces used in Part 1. In the *disgust* condition, participants viewed a 10s video consisting of a sequence of 10 disgust pictures from the IAPS. In the *fear* condition, participants viewed a 10s video of 10 fear pictures from the IAPS. In the control condition, participants viewed a 10s video consisting of short clips of human or (not uncanny) robots used in Part 1. Participants were randomly assigned to one of the four conditions, watched the assigned video, and immediately continued with the lexical decision task.

The lexical decision task was adapted from Huang and Wyer (2015). Participants had to decide whether the letter string presented is a real word or not by pressing one of two buttons corresponding to real or not real. Response times were recorded for the three different categories of words. Word presentation order was randomized.

In the final rating and categorization task, participants were presented with the still body stimuli and were asked to rate the stimuli on the scales used in Part 1. After the ratings, participants were presented with a two-alternative forced-choice task with the categories *normal* and *not normal*. Participants had an unlimited amount of time to respond. Body stimuli were presented in random order.

Data analysis and availability. Data preparation and analysis was performed using R version 4.1.2 and JASP. Mixed-effects models were used given the between-within subject design of the experiment. For R, the packages lme4, lmer, and lmerTest were used (Bates, Mächler,

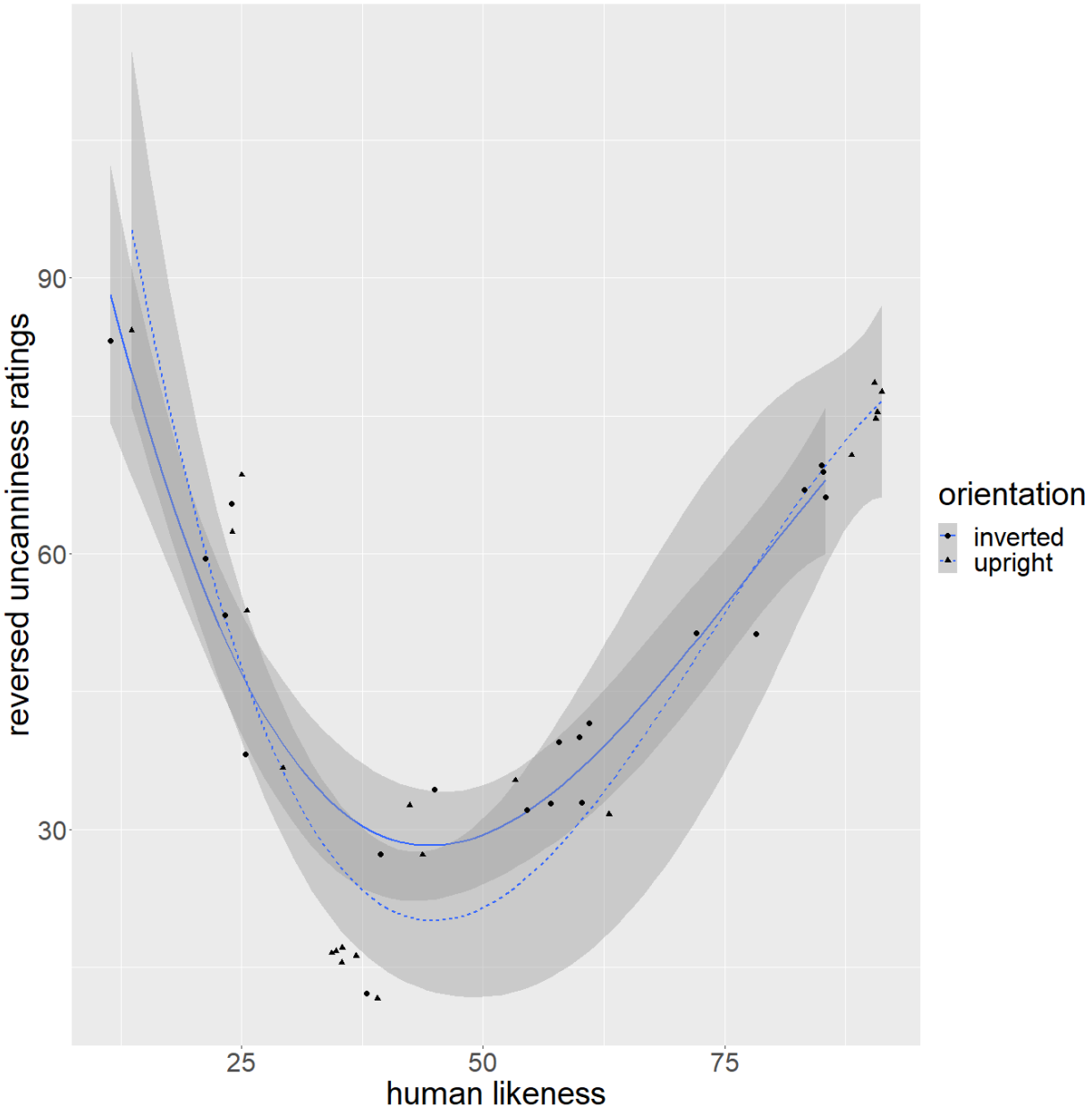
Bolker, & Walker, 2015). For the priming analysis, lexical decision task trials with incorrect word classification were excluded, just as reaction time outliers. The data, analysis scripts, and body stimuli are available online at https://osf.io/zynf7. This study's design and its analysis were not pre-registered.

Results and Discussion

Part 1: Uncanny valley and orientation. Outlier removal was conducted on a by-stimulus level for both uncanniness and human likeness ratings. A total of 17 uncanniness outliers and 25 human likeness outlier values were removed. The effect of orientation on the uncanny valley was analysed by computing a linear mixed model with orientation interacting with a linear, quadratic, and cubic function of human likeness as fixed factors, and participants and stimuli as random factors. Results reveal that a non-linear function of human likeness could explain uncanniness ($t(4883) = 5.07$, $p < .001$), and that orientation also interacted with linear ($t(4855) = 2.35$, $p = .019$), quadratic ($t(4852) = 2.51$, $p = .012$), and cubic ($t(4855) = 2.19$, $p = .028$) human likeness ($R^2_{cor} = .61$). *Figure 8.1* shows the interaction between orientation and human likeness. As an uncanny valley-like function could be plotted for both upright and inverted conditions, the configural processing hypothesis 1 was only partially supported.

Figure 8.1

Cubic lines of estimate for upright and inverted video stimuli. Gray areas indicate standard errors.



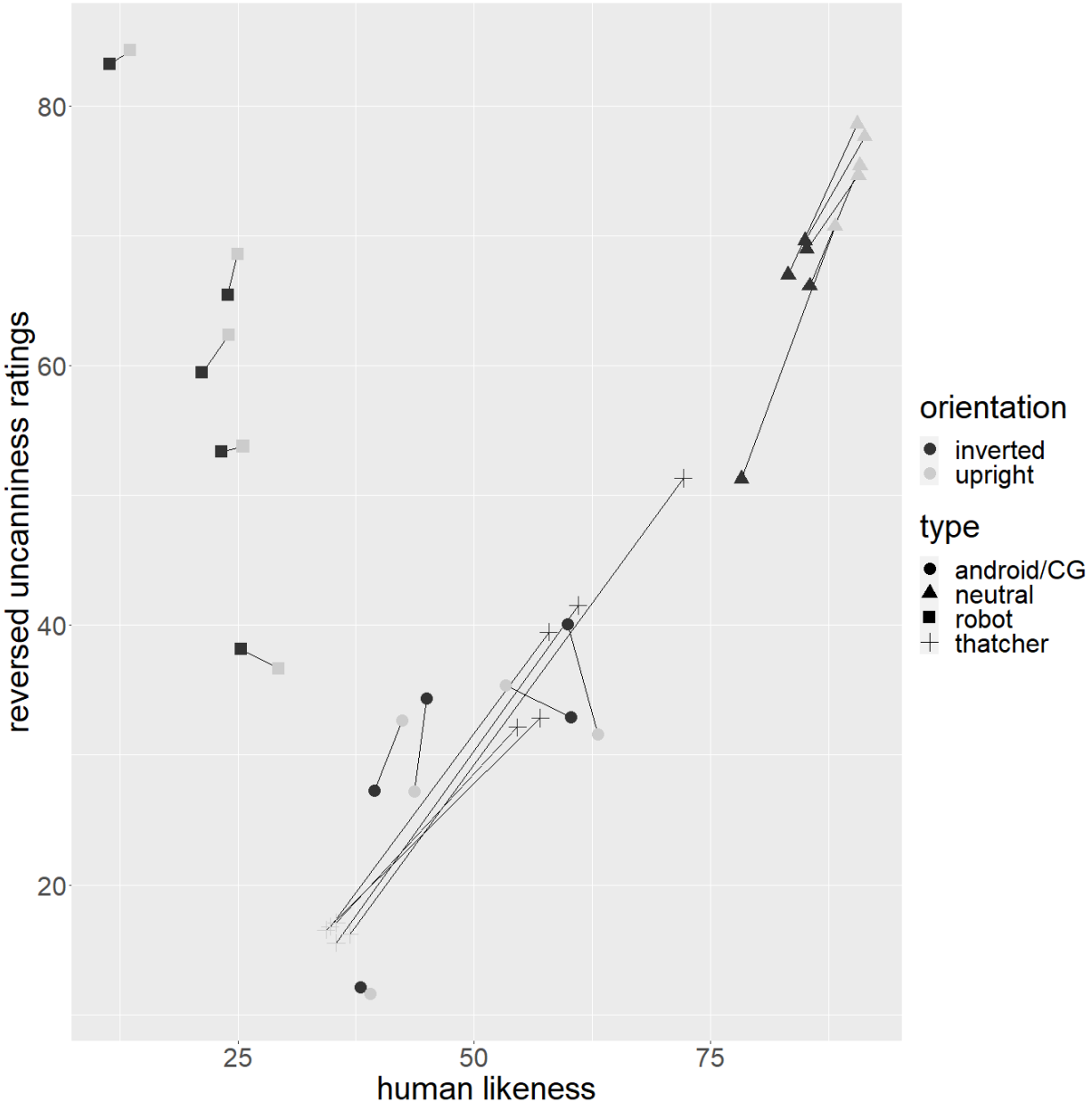
To further investigate the effect of orientation on uncanniness and how it differs between types of stimuli, a linear mixed model has been calculated with orientation and type as fixed effects and participants and stimuli as random effects. The model (R^2_{cor} = .59) shows a significant main effect of type (t(337) = 38.33, p < .001) and an interaction between type and orientation (t(332) = 4.34, p = .011), but no main effect of orientation (t(332) = 0.89, p

= .353). Post-hoc Tukey tests show that while inversion increased the uncanniness of normal humans ($t(4862) = -8.26$, $p_{adj} < .001$, $d = 1.29$), it decreased the uncanniness of Thatcher humans

($t(4862) = 18.17$, $p_{adj} < .001$, $d = 2.84$). However, inversion did not affect robots ($t(4862) = 1.09$, $p_{adj} = .69$) or androids ($t(4862) = -1.27$, $p_{adj} = .610$). The data is plotted by stimulus in *Figure 8.2* on a stimulus level.

Figure 8.2

Mean uncanniness and human likeness ratings of upright and inverted stimuli divided by stimulus type.

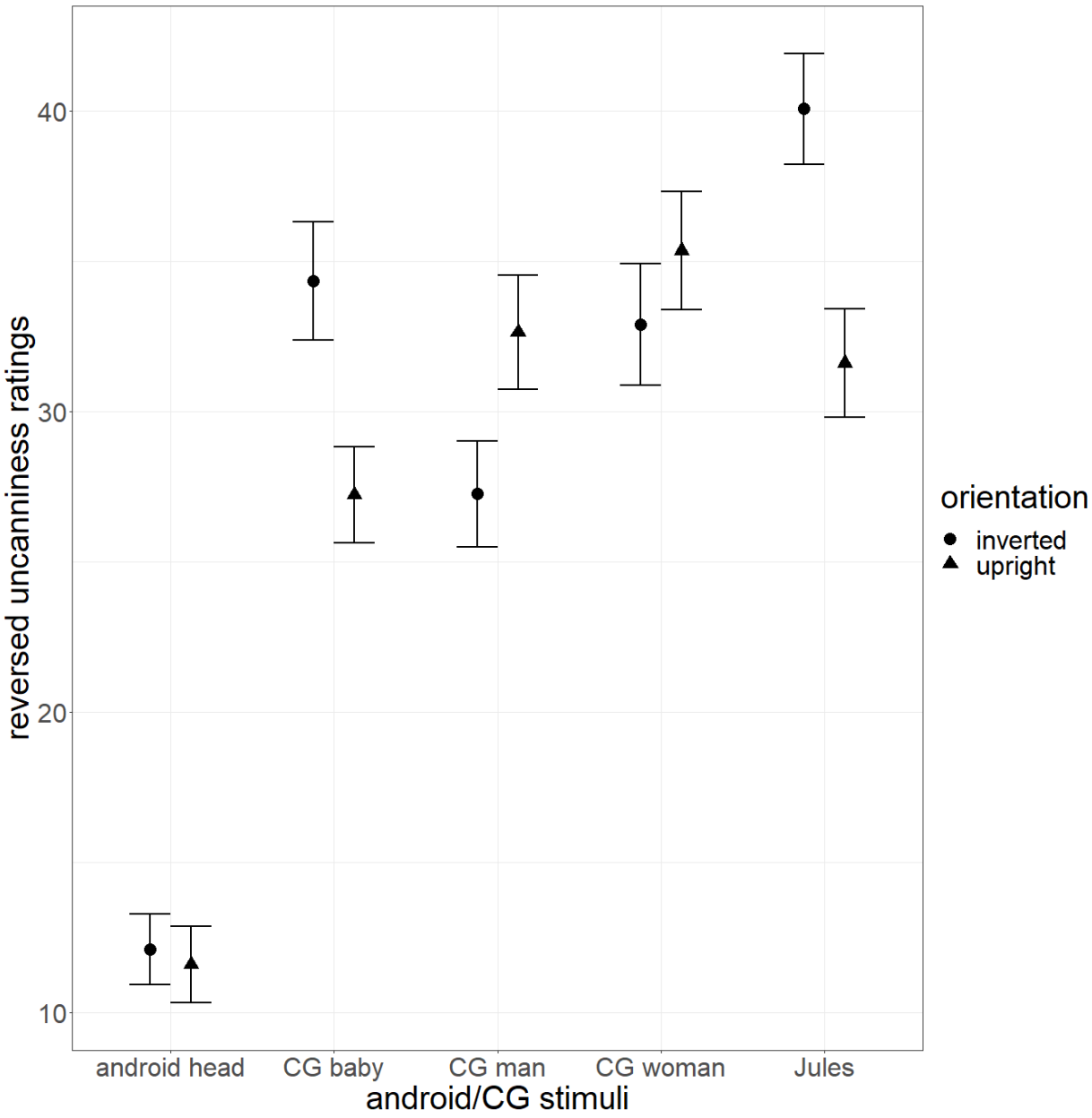


The data depicted in *Figure 8.2* indicates that within the android condition, orientation either increased or decreased uncanniness depending on the specific stimulus. The data was closer plotted in *Figure 8.3*, and an additional linear mixed model with base stimuli and orientation as fixed effects and participants as random effects was performed to investigate the effect of orientation on the stimulus level. The model (R^2_{cor} = .42) shows a significant main effect of base stimulus ($t(4133) = 72.42$, $p < .001$) and a significant interaction between base stimulus

and orientation ($t(4633) = 11.58$, $p < .001$), but no main orientation effect ($t(1159) = 358$, $p = .060$). Post-hoc Tukey tests further revealed that inversion reduced the uncanniness of the android Jules ($t(1111) = 3.38$, $p_{adj} = .001$, $d = 0.43$) and the CG baby (Tin Toy; $t(1111) = 3.31$, $p_{adj} = .001$, $d = 42$), but increased the uncanniness of the CG man (Apology; $t(1111) = 2.58$, $p_{adj} = .010$, $d = 0.33$). Inversion did not affect the uncanniness of the CG woman (Mary Smith; $t(1111) = –1.17$, $p = .243$) or the android head ($t(1111) = 0.33$, $p_{adj} = 1.00$). Thus, inversion—a proxy for configural information—is relevant to the uncanniness of only some androids or CG characters. Because an effect of inversion was observed only for Thatcher faces and some of the android/CG stimuli, configural processing *hypothesis 2* was only partially supported.

Figure 8.3

Uncanniness ratings for individual android and CG stimuli, both upright and inverted. Error bars indicate standard error.



The investigation into the role of inversion as a proxy for configural processing in its role in the uncanny valley revealed only a partial importance of configural processing. While an uncanny valley like function (*Figure 8.1*) seemed to be slightly flattened in the inversion condition, it was nevertheless present, indicating that certain uncanny visual information survives a disruption of configural processing. Similarly, inversion only reduced the

uncanniness of some android/CG stimuli, namely the android Jules and the CG baby, but did not affect or increased the uncanniness of other uncanny stimuli. Thus, the role of configural processing on uncanniness seems to depend on the specific stimulus. In other words, configural information plays a role in uncanniness only in some instances.

Previous research on the role of configural information on uncanniness showed that configural distortions are still uncanny in an inverted face, but that the increase of uncanniness with increasing distortion is reduced more than in upright faces (Chapter 2). Thus, certain configural information may still survive inversion. In this sense, failing to find a complete elimination of uncanniness through inversion does not rule out the role of configural information in the uncanny valley. However, the finding that the uncanny valley curve is "flattened" through inversion does support previous research on the role of configural information on aesthetics ratings (e.g., Leder et al., 2017; Santos & Young, 2008). Various errors may occur during the design of an artificial humanlike entity, which may be on a featural level or on a feature-relational level. In some instances, like the android Jules, deviations on the configural level specifically could be a source of uncanniness, while in other instances, distortions in individual features, mismatched features (Seyama & Nagayama, 2007), missing features (e.g., the missing body of the android head in this study), or other design issues may be the source of uncanniness. Furthermore, certain configural information may remain intact after inversion: As configural processing depends on experience with specific patterns of motion (Wang et al., 2022), inversion-invariant motion configurations may survive an inversion. Thus, inversion may not completely disrupt configural processing. In general, however, the observed results support the notion that an uncanny valley can be caused by multiple mechanisms, or by different dimensions on which those mechanisms could become relevant (e.g., mismatches in face vs. body perception). The relevance of each explanation would then depend on the individual uncanny stimulus (Diel &

MacDorman, 2021; Kim, de Visser, & Phillips, 2022; Strait et al., 2017). However, these explanations are ad hoc and different, stimulus-dependent causes of uncanniness require further empirical verification.

Part 2: Lexical Decision Task. Outlier removal was conducted on a by-stimulus level for reaction times for each prime condition. A total of 46 outlier values were removed. To compare the effect of uncanniness priming on semantic associations with disgust and fear priming, a mixed-design ANOVA with condition as a between-subject variable and word type as a within-subject variable was calculated. Results reveal no main effects of condition ($F(3, 124) = 0.18$, $p = .909$) or word type ($F(2, 248) = 2.51$, $p = .084$), but a significant interaction ($F(6.248) = 2.68$, $p = .015$, $\eta^2 = .026$).

Post-hoc Tukey tests showed significant differences between conditions: In the control priming condition, reaction times for mortality-related words were not higher than normal words ($t(2720) = 0.22$, $p_{adj} = 1.00$), nor were disease-related words ($t(2720) = –0.69$, $p_{adj} = .488$). In the disgust priming condition, however, disease-related words had longer reaction times than normal words ($t(2720) = 3.38$, $p_{adj} < .001$, $d = 0.87$), while mortality-related words did not ($t(2720) = –0.45$, $p_{adj} = 1.00$). In the fear priming condition, the opposite was the case: disgust-related words did not have a longer reaction time than normal words

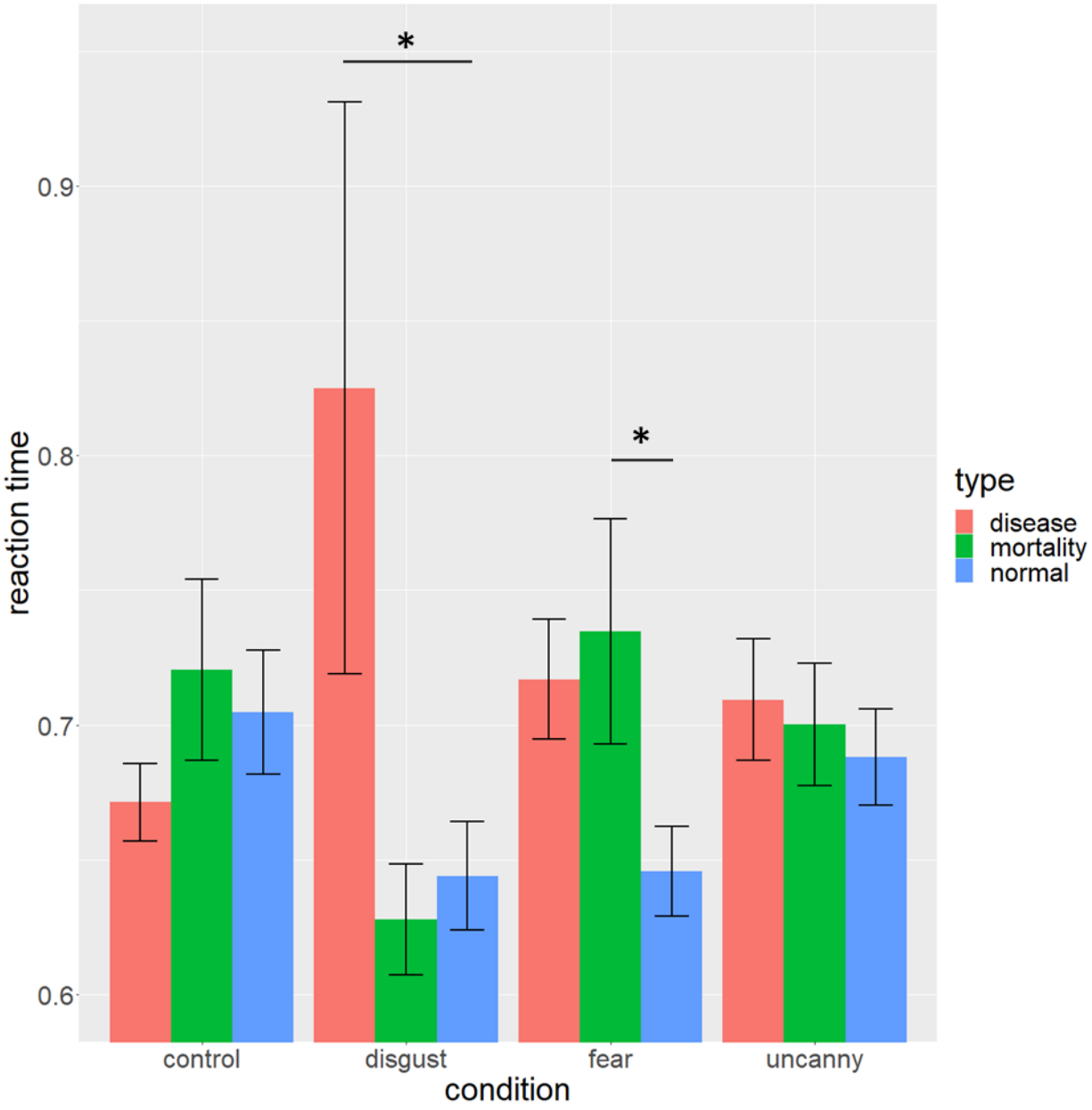
($t(2720) = 1.28$, $p_{adj} = .100$), but fear-related words did ($t(2720) = 1.74$, $p_{adj} = .041$, $d = 0.45$).

Finally, in the uncanny priming condition, neither disease-related words ($t(2720) = 0.33$, $p_{adj} = 1.00$) nor mortality-related words ($t(2720) = 0.1$, $p_{adj} = 1.00$) had longer reaction times than normal words. Thus, while successful priming effects of disgust and fear stimuli have been observed on disease-related and death-related concepts, such effects were not observed for uncanny primes. The data is summarized in *Figure 8.4*. Because uncanniness primes affected

the processing of disease- or death-related words while disgust- or fear primes did, the death priming hypotheses 1 and 2 and disease priming hypotheses 1 and 2 were not supported.

Figure 8.4

Mean reaction times divided by priming conditions and word types. Error bars indicate standard errors. Asterisks mark tested significant differences.



While the reaction times of disease- or death-related words did not differ in the control prime condition, disgust primes increased reaction times for disease-related words but not for death-related words compared with neutral words, while fear primes increased reaction times for death-related words but not for disease-related words compared with neutral words. In addition, a difference is implied in *Figure 8.4* between normal and disease-related words in the fear condition, although Bonferroni-adjusted p-values did not show a significant difference. Increased reaction times here may reflect a priming of avoidance behaviour targeted at emotion-specific stimuli (i.e., avoidance of disease-related concepts for disgust primes, avoidance of death-related concepts for fear primes). An uncanny prime, however, did not affect the reaction times of disease- or death-related words compared with neutral words. As the increase of reaction times can be seen as an indicator of successful priming of either disease or death concepts, the results indicate that uncanny stimuli do not activate such concepts, contrary to the predictions of disease avoidance and mortality salience theories.

The results indicate that the uncanny valley observed in this study is not associated with disease avoidance or mortality salience. These results contradict previous research, for example findings of associations between disgust or disgust sensitivity and the uncanny valley (Ho et al., 2008; MacDorman & Entezari, 2015). It is possible that certain features in an uncanny entity elicit a certain measure of disgust (e.g., distorted body parts or motions as indicators of disease; Park, Faulkner, & Schaller, 2003). However, those features may not be the main source of uncanniness. As uncanniness can be observed with stimulus types that do not have a clear danger of disease contamination (Chapters 5 and 6; Diel & MacDorman, 2021; Freud, 1917), uncanny android stimuli may also elicit uncanniness through cognitive mechanisms unrelated to the avoidance of disease, but instead, for example, violations of

expectations (Saygin, Chaminade, Ishiguro, Driver, & Frith, 2012) or deviations from experienced norms (Chapters 2 to 4).

The present results contradict previous research associating the uncanny valley with mortality salience (Koschate et al., 2016; MacDorman, 2005). Koschate et al. (2016) found that uncanny androids increase death-thought accessibility, and MacDorman (2005) found that uncanny androids increased cognitive biases associated with terror management theory, such as support for nationalistic leaders. However, anxiety-inducing stimuli may trigger death-related thoughts or cognitive biases without directly reminding the viewer of their own mortality (e.g., by resembling a dead human body). In fact, the present study found that a general fear prime (e.g., snakes, spiders) affects the reaction times of death-related words, even though general fear has not been associated with mortality salience in past research. Thus, the uncanniness of androids (which is associated with fear; Ho et al., 2008) could have caused the results in previous research but may not have been strong enough to affect death-related words in the present research. In any case, the current results do not support mortality salience as an explanation of the uncanny valley. Similarly, danger avoidance, at least when caused by the perception of a dead body (Moosa & Ud-Dean, 2010), is also not supported by the present results.

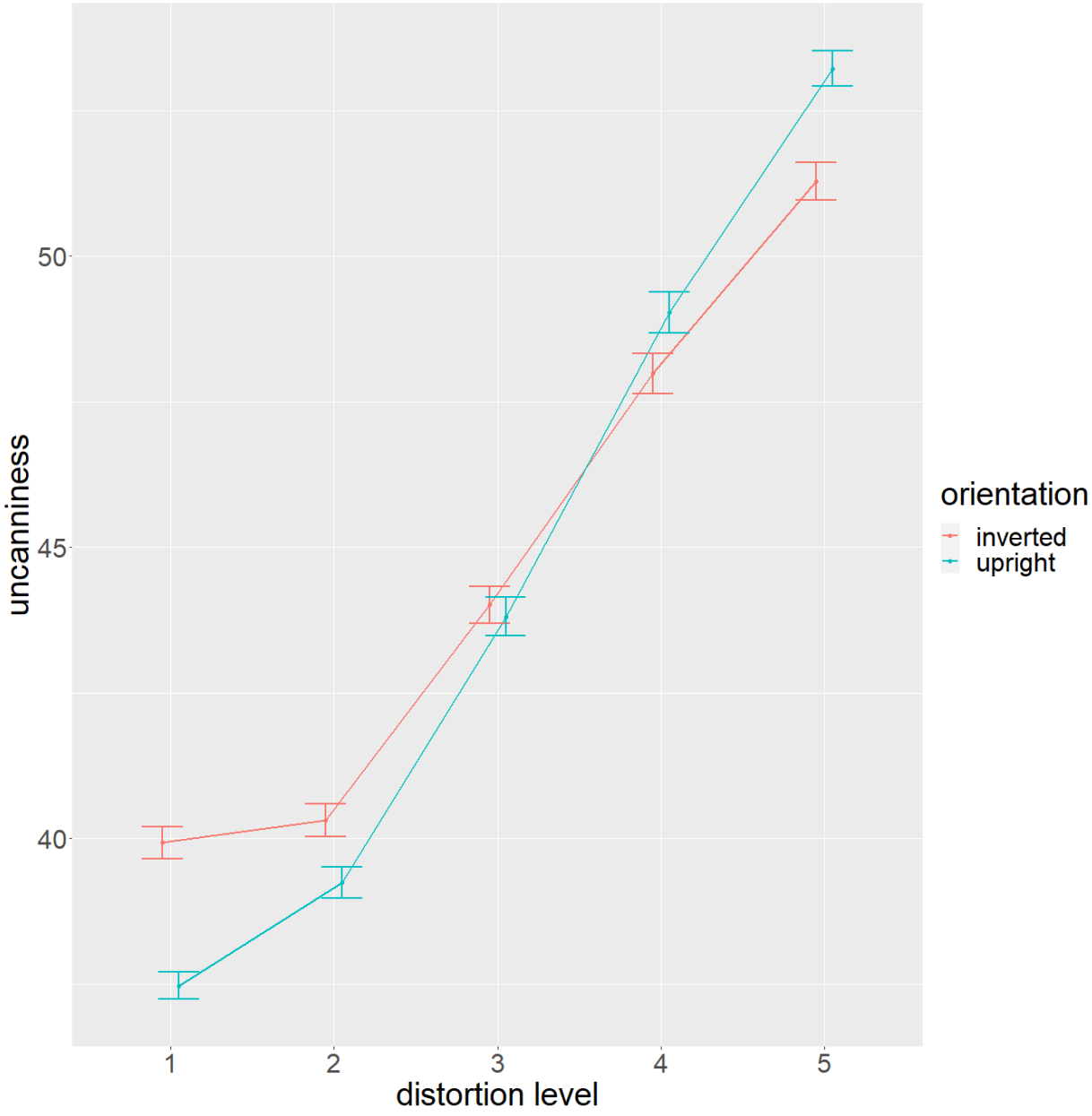
However, no systematic approach has been used to collect the words for each condition. Thus, while significant results were found, it is unclear whether the used words correspond to their respective concepts. In addition, the stimuli in both the control and uncanny prime conditions were shown in the previous task which may have influenced the priming procedure. Repetitive presentation of emotional stimuli may lead to habituation, decreasing physiological and emotional responses towards the stimulus (Foa & Kozak, 1986). Habituation to uncanny androids may have decreased potential priming effects in the uncanny condition.

Part 3: Body rating and categorization. Outlier removal was conducted on a by-stimulus level for uncanniness ratings and categorization reaction times. A total of 83 uncanniness and no reaction time outlier values were removed. For body rating, linear mixed models were calculated with orientation and distortion level as fixed effects and participants and stimuli as random effects. The model (R^2_{cor} = .92) shows a significant main effect of distortion level (t(4228) = 137.95, p < .001) and an interaction between orientation and distortion level (t(41252) = 18.04, p < .001), but no main effect of orientation (t(1133) = 0.4, p = .528). Post-hoc Tukey tests show that at distortion level 1 (no distortion), inverted bodies were more uncanny than upright bodies (t(10404) = 5.99, p_{adj} < .001, d = 0.72). At distortion level 2, inverted bodies were also more uncanny than upright bodies (t(10404) = –2.55, p_{adj} = .016, d = 0.31). At distortion level 3, inverted bodies were no longer more uncanny (t(10404) = –0.49, p_{adj} = .933), and at distortion level 4 (t(10404) = –5.99, p_{adj} < .001, d = 0.3) and level 5 (t(10404) = –5.99, p_{adj} < .001, d = 0.57), upright bodies were more uncanny than inverted bodies. Thus, while inverted bodies were more uncanny than upright bodies at lower distortion levels, upright bodies were more uncanny than inverted bodies at higher distortion levels. The data are depicted in *Figure 8.5.*

Because distortions increased the bodies' uncanniness ratings and categorizations as not normal, and the effect was stronger in the upright condition than the inverted condition, configural processing hypotheses 3 and 4 are supported.

Figure 8.5

Mean uncanniness ratings of bodies across distortion levels, both upright and inverted. Error bars indicate standard errors.

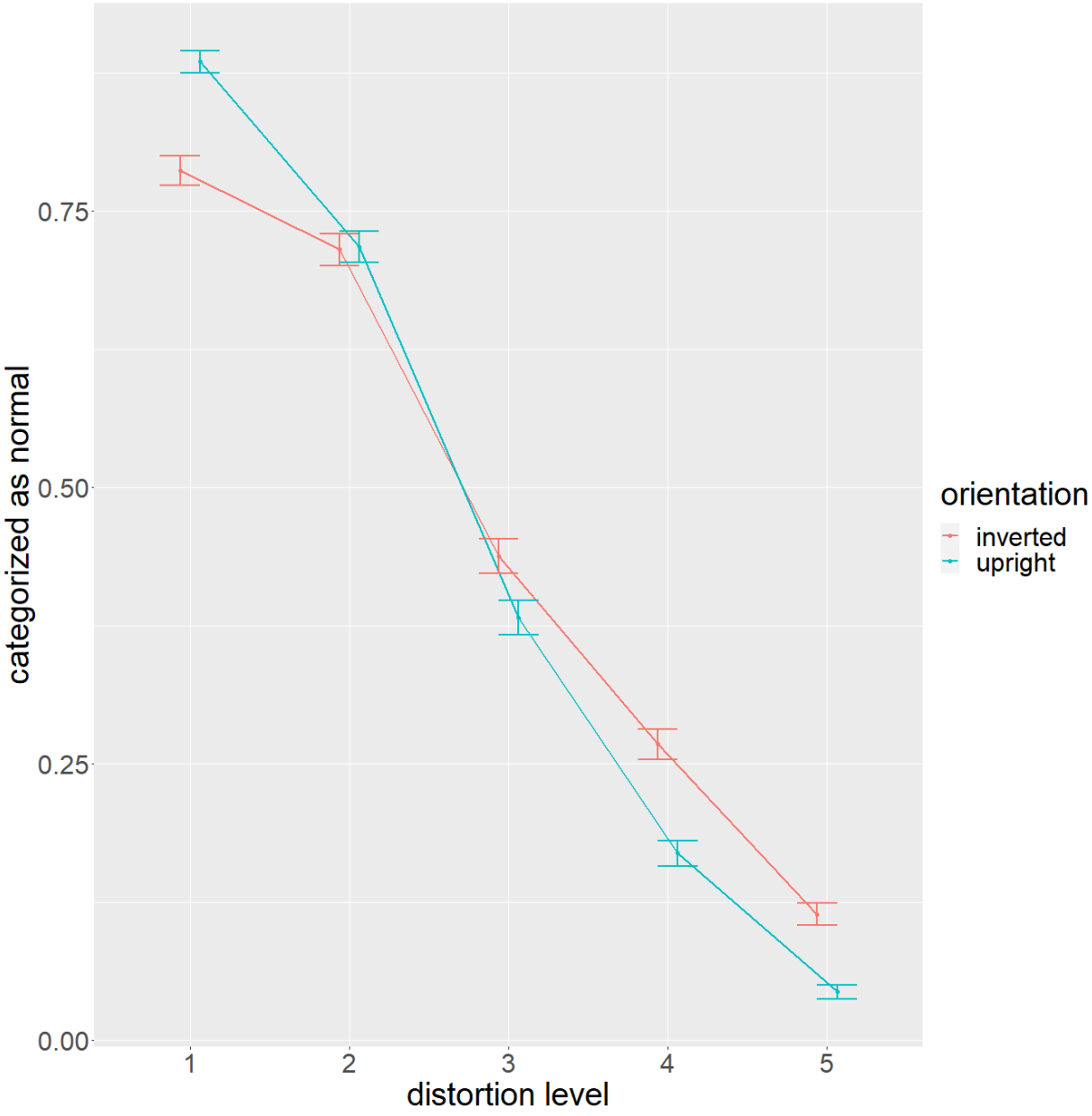


Body categorization data showed results similar to uncanniness ratings: Linear mixed models with orientation and distortion level as fixed factors and participant as a random factor (R^2_{cor} = .92) showed significant main effects of orientation ($t(1149) = 7.44$, $p = .007$), distortion level ($t(4298) = 1090.89$, $p < .001$), and an interaction ($t(4903) = 20.86$, $p < .001$). Post-hoc Tukey tests show that while inverted bodies were less likely to be categorized than normal bodies at distortion level 1 (no distortion; $t(10404) = 5.76$, $p_{adj} < .001$, $d = 0.71$), there was no

difference at distortion level 2 ($t(10404) = 0.09$, $p_{adj} = .929$). However, inverted bodies were more likely than upright bodies to be categorized as normal at distortion level 3 ($t(10404) = -3.17$, $p_{adj} = .005$, $d = 0.4$), level 4 ($t(10404) = -5.74$, $p_{adj} < .001$, $d = 0.71$), and level 5 ($t(10404) = -4.13$, $p_{adj} < .001$, $d = 0.5$). Thus, inversion reduced the tendency to categorize distorted bodies as not normal especially at higher distortion levels. The data are depicted in *Figure 8.6*.

Figure 8.6

Mean percentages of categorizations as "normal", across distortion levels, both upright and inverted. Error bars indicate standard errors.



The present results mirror previous research on the uncanniness of, and ability to detect, deviations of human faces (Chapter 2): Stronger deviations from the typical appearance appear uncanny and are more likely to be considered not normal, although these effects are reduced when stimuli are presented inverted, indicating the role of configural processing in the detection of deviations. The current results indicate the role of configural processing in

the uncanny valley may extend beyond faces (Chapter 2) to full bodies, when deviations or mismatches occur in the configuration of body parts.

Experiment 13

While the results in Experiment 12 are promising, the interpretation of the priming results are limited by the non-systematic choice of target words. In addition, priming stimuli in the uncanny and control conditions were shown in the task prior to the priming, which may influence (e.g., weaken) priming effects. To validate the results and replicate them using a validated set of target words, Experiment 13 has been conducted to replicate Part 2 of Experiment 12 using a new, validated set of target words.

Hypotheses

The hypotheses of Experiment 13 are identical to those of the priming study in Experiment 1:

- *Death priming hypothesis 1*: Uncanny primes change reaction times for death-related words in a lexical decision task
    - *Death priming hypothesis 2*: Reaction times for death-related words do not differ between uncanny primes and fear primes.
- *Disease priming hypothesis 1*: Uncanny primes change reaction times for disgust-related words in a lexical decision task
    - Disease priming hypothesis 2: Reaction times for disease-related words do not differ between uncanny primes and disgust primes.

Methods

Participants. Uncanny primes in the previous Experiment may have been too weak compared with the (highly arousing) disgust and fear primes to elicit stronger effects. Because no previous research on uncanny priming exists, a power analysis has been calculated using a

standard small effect size of $d = 0.25$. Using a 2 × 3 mixed model, power analysis revealed that 89 participants per condition (n = 356) would be sufficient to reach a power of 0.8. Participants were selected from the same pool as in Experiment 1, but did not take part in Experiment 1 or the pilot studies of Experiment 2. Participants' average age was $M_{age} = 19.37$ ($SD_{age} = 1.82$), 305 were female, 39 male, and 15 preferred not to answer.

Target word validation. To validate the semantic association of the target words, two pilot studies have been conducted to select a set of target words highly associated with the conditions (disease, mortality, normal/control, nonsense word). Analogous empirical pilot studies for word stimulus selection have been used in previous research involving lexical decision tasks (e.g., Rossell & Nobre, 2004; Yao & Wang, 2013).

In the first control study, 26 participants were asked to come up with as many disease- or mortality-related words as possible. All words mentioned by at least two participants were then selected as stimuli for the second control study. The words can be found on OSF.

In the second control study, participants conducted a four-choice forced categorization task with the new set of words, also including an extended list of neutral control words and made-up letter strings. Participants had to categorize each word as quickly as possible as one of the following four semantic categories: disease, mortality, neutral (a real English word neither associated with disease or mortality), or nonsense (not a real English word). After averaging across participants, eight words with the highest categorization consistency were selected for each condition. These words are summarized in *Table 3*. To investigate potential differences in word length of word frequency between the condition, two ANOVAs. However, no effects of conditions have been observed for either word length ($F(2,21) = 0.012$, $p = .989$) or frequency ($F(2,21) = 1.03$, $p = .374$). Thus, word length and frequency are comparable across conditions.

Table 8.2

Words used in Experiment 2, divided by condition (word type).

Neutral words	Disease-related words	Death-related words	Non-words
Bacon; book; English; grass; milestone; miniature; park; teacher	Cure; germs; Infection; illness; medicine; sick; symptoms; virus	Coffin; grave; graveyard; heaven; hell; killed; mourning; skull	Afer; dopleek; delliv; falipi; falgo; fathis; glamasaka; groleht; grusdi; holdok; horrk; kininal; krable; midaun; roqua; semnp; sgaal; solos; suggry; talal; tlook; tybs; wrinbel

All disease- and death-related words used in Experiment 1 were above 75% of consistency for their respective conditions, aside from nausea (disease; 70%), rash (disease; 63%), doom (death; 55%), and pest (disease; 44%). Validity of the death-related and control target words are thus affirmed, and partially affirmed for disease-related words. Nevertheless, a replication has been conducted using only the most consistent target words summarized in *Table 3*.

Prime stimuli. Disgust and fear primes in Experiment 2 were identical to the primes used in Experiment 1. However, because using picture-based stimuli for disgust and fear primes and

video-based stimuli for control and uncanny primes was a potential confounding variable in Experiment 1, pictures of the uncanny androids/CG characters and not uncanny robots/humans from Experiment 1 were used instead of videos. Thus, all priming material was picture-based but otherwise remained identical to Experiment 1.

Procedure. The procedure of Experiment 1 was identical to Part 2 of Experiment 1, with the exception of the study being conducted online. Online research can reliably replicate lab-based research except for potential additive reaction time effects (Reimers & Stewart, 2015; Semmelmann & Weigelt, 2017), including priming tasks (Angele, Baciero, Gómez, & Perea, 2023).

Data analysis and availability. Data preparation and analysis were performed using R version 4.1.2 and JASP. Mixed-effects models were used given the between-within subject design of the experiment. For R, the packages lme4, lmer, and lmerTest were used (Bates, Mächler, Bolker, & Walker, 2015). For the priming analysis, lexical decision task trials with incorrect word classification were excluded, just as reaction time outliers. The data, analysis, and body stimuli are available online at https://osf.io/zynf7. This study's design and its analysis were not pre-registered.

Results and Discussion

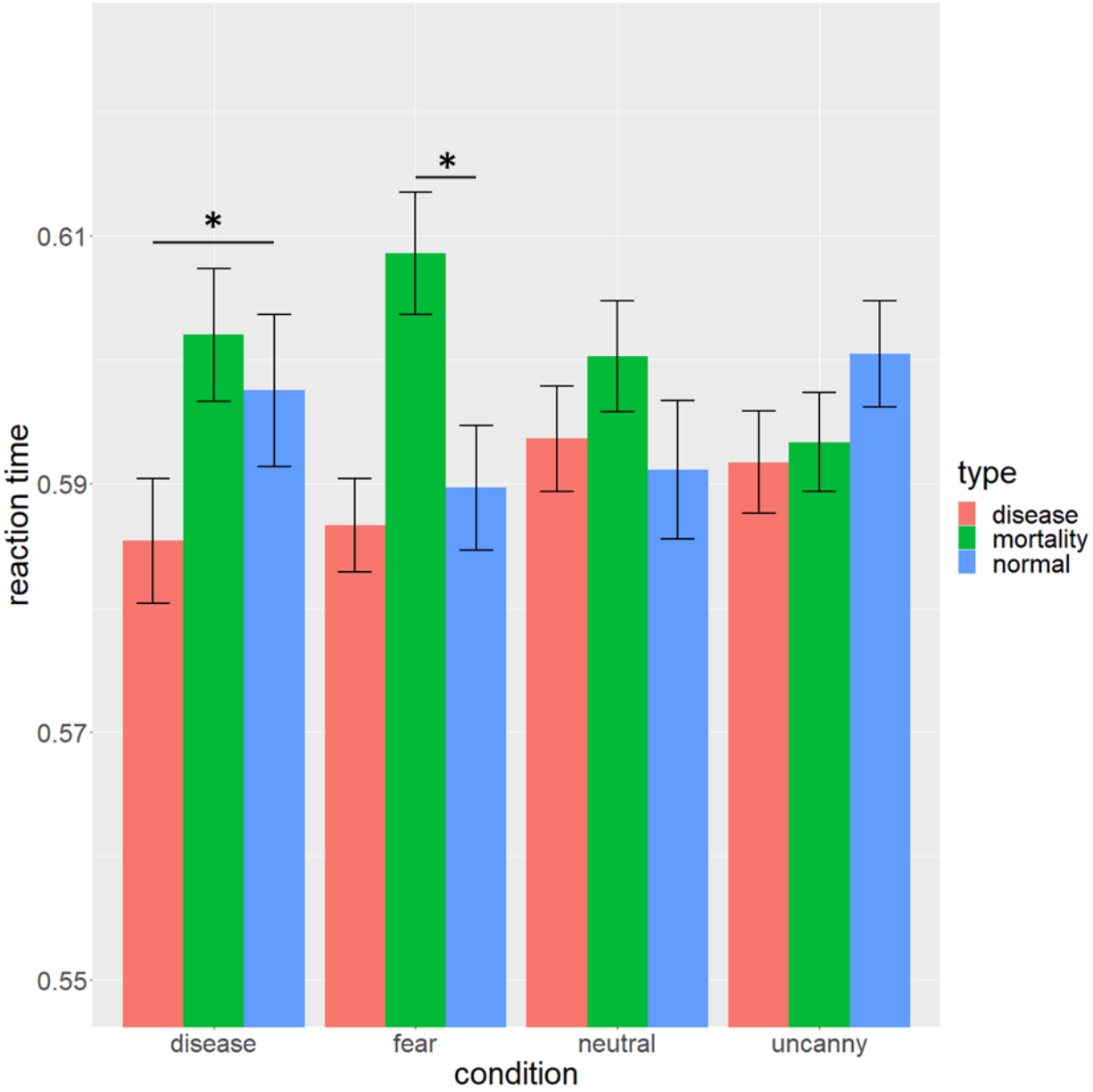
A total of 24 outlier values were removed. To compare the effect of uncanniness priming on semantic associations to disgust and fear priming, a mixed-design ANOVA with condition as a between-subject variable and word type as a within-subject variable was calculated, with stimuli and participants as error terms. a significant interaction ($F(6,7066) = 2.48$, $p = .002$, $\eta^2 = .002$).

Post-hoc comparisons revealed that reaction times were lower for disease-related words than neutral words following disease primes ($t(6400) = -1.81$, $p_{adj} = .035$, $d = 0.56$), but not

following fear ($t(6400) = -0.49$, $p_{adj} = .314$), uncanny ($t(6400) = 0.511$, $p_{adj} = 0.66$), or control primes ($t(6400) = -1.44$, $p_{adj} = .074$). Reaction times were increased for mortality-related words following a fear prime ($t(6400) = 3.94$, $p_{adj} < .001$, $d = 1.16$), but not a disgust ($t(6400) = 0.04$, $p_{adj} = .485$), uncanny ($t(6400) = -1.56$, $p = .941$), or control prime ($t(6400) = -1.41$, $p_{adj} = .079$). Thus, disgust primes lowered reaction times for disease-related words, while fear primes increased reaction times for mortality-related words. These effects, however, were not observed for uncanny or control primes. The data is summarized in *Figure 8.7*.

Figure 8.7

Mean reaction times divided by priming conditions and word types. Error bars indicate standard errors. Asterisks mark tested significant differences.



Interpretation of results

Similar to Experiment 1, disgust and fear primes affected reaction times for disease- and death-related words, respectively. Uncanny and control primes did not. However, in this replication, disgust primes *decreased* reaction times for disease-related words, whereas in Experiment 1, reaction times were increased. Meanwhile, fear primes increased reaction times for death-related words in both experiments.

Given that uncanny stimuli may be uncanny for different reasons, only some of the uncanny primes used here may have had disease-avoidance or mortality-salience related effects, while others may have even counteracted those. Given the possible heterogeneity of the effect, further research is needed to investigate the effects of uncanny primes on a stimulus-level basis. Nevertheless, the results contradict disease avoidance or mortality salience as general explanations of the uncanniness of stimuli falling into an uncanny valley.

The discrepancy in results between Experiment 1 and 2 are interesting: Although both experiments found priming effects caused by disgust stimuli, the results go in opposites directions. Different semantic stimuli may have caused this discrepancy. Most disease-related words in Experiment 1 were directly related to diseases and symptoms, which could trigger automatic vigilance since negative information holds attention longer, increasing reaction time in the process (Estes & Adelman, 2008; Yao et al., 2019). In Experiment 2 however, some disease-related words were related to the treatment of disease (cure, medicine). If automatic vigilance increases the reaction time of disease-related words following a disease prime, the effect may not affect treatment-related words because they do not hold negative emotional content. Instead, disease primes may even prepare approach behaviour towards treatment-related information reflecting an adaptive strategy towards indicators of disease, and decreasing reaction times in the process. Similarly, the discrepant results between fear and disgust primes on mortality and disease words (increased reaction times for mortality words after fear primes; decreased reaction times for disease words after disgust primes) may reflect different effects of target words: While some disease-related words may have activated approach-related behaviour, mortality-related words may have consistently elicited mechanisms increasing reaction time (Yao et al., 2019). In any case, the discrepancies between the results in Experiment 1 and 2 pose difficulties in interpreting the consequences of disgust primes. Nevertheless, as the pattern elicited by disease primes was not found for

uncanny primes in neither experiment, the results do not support a disease avoidance explanation of the uncanny valley.

Increased reaction times for mortality-related words following a fear prime may reflect defence mechanisms following a reminder of death. Terror management theory postulates that reminders of mortality elicit defence mechanisms like the suppression of death-related thoughts or focussing more on death-unrelated information (Greensberg et al., 1990; Hayes, Schimel, Arndt, & Faucher, 2010). Suppressed processing of mortality-related content (indicated by increased reaction times for death-related words) may thus reflect such a defence mechanism. Alternatively, increased reaction times for death-related words following fear primes may reflect automatic vigilance processes described above. As uncanny primes did not show these effects, the results do not support the explanation that the uncanny valley observed in this study is associated with mortality salience.

In Experiment 1, control and uncanny prime stimuli were presented in the task prior to the priming which may have weakened the effects of prime stimuli. However, the same effects of control and uncanny primes were found here in Experiment 2. The effects on the control and uncanny primes cannot be explained by repeated exposure to the prime stimuli.

General Discussion

The goal of the present work was to critically investigate different explanations on the uncanny valley. The current work was the first to investigate the effect of inversion on uncanniness using uncanny stimuli typically associated with the uncanny valley (e.g., androids and robots). Partial support for the configural processing explanation was found: Inversion reduced uncanniness in some, but not all, android/CG stimuli. However, the effect was not comparable to the effect of inversion in the Thatcher faces. In addition, the findings of Chapter 2 were replicated using body instead of face stimuli: Increasing distortions from

typical appearance are rated as more uncanny, but less so when bodies are presented inverted, indicating that configural information is used for the assessment of uncanniness caused by deviations. Thus, while configural information can be relevant for the uncanniness of deviating face or body stimuli, this mechanism seems to be relevant only for some instances of the uncanny valley.

This work is also the first to find evidence against two common theories of the uncanny valley: disease avoidance and mortality salience. Uncanny stimuli used as primes did not have the same effects as disgust or fear prime stimuli, which affected reaction times to disease- or death-related words. Thus, stimuli falling into the uncanny valley do not seem to be conceptually related to disease and death, which would be expected from theories on disease avoidance or mortality salience, or danger avoidance in relation to dead bodies.

While previous research found consistent support for the refined theory, its underlying neural mechansism remain unclear. Two neurocognitive theories may best explain the processes proposed by the refined theory: disfluency caused by the processing of deviating stimuli (Winkielman et al., 2003), or prediction errors elicited by the mismatch between the (deviating) sensory input and the expectation (Flogel, 2010). Chapter 9 will investigate these explanations.

Chapter 9: Electrophysiological correlates of face processing and prediction error and the uncanny valley

Methods, experiment, and large portions of the introduction and discussion in this chapter is currently in review in the journal *Neuropsychologia*.

Introduction

While the previous chapters found evidence of cognitive effects related to specialized processing, evidence of neural correlates remains lacking. If a specialized processing account is correct, then increased neural activity markers of specialized processing (e.g., face-related event-related components) would support the role of specialized processing in the uncanny valley.

Neural correlates of face processing

Links between face uncanniness and face-sensitive processing should be reflected in neural correlates of face processing. The *fusiform face area* (FFA) in the middle fusiform gyrus is sensitive to configural information in faces compared to other stimulus categories (Kanwisher & Moscovitch, 2000). The N170 component, a negative event-related potential (ERP) approximately 150-200 milliseconds after stimulus onset, has also been associated with configural face processing (Eimer, 2011; Olivares, Iglesias, Saavedra, Trujillo-Barreto, & Valdés-Sosa, 2015) and is estimated to have its source at the FFA (Olivares, Lage-Castellanos, Bobes, & Iglesias, 2018). Finally, the P100 component precedes the N170 component in face processing and is thought to correlate with earlier stages of more feature-based processing (Herrmann, Ehlis, Ellgring, & Fallgatter, 2004).

Deviating faces elicit stronger FFA activity and delayed and increased N170 components compared to typical faces, as long as the global configuration remains intact (Carbon, Schweinberger, Kaufmann, & Leder, 2005; Cassia, Kuefner, Westerlund, & Nelson, 2006;

Hahn, Jantzen, & Symons, 2012; Halit, de Haan, & Johnson, 2003; Loffler, Yourganov, Wilkinson, & Wilson, 2005; Mattevelli et al., 2013; Milivojevic, Clapp, Johnson, & Corballis, 2003; Said, Dotsch, & Todorov, 2010; Workman et al., 2021). An increased activity for deviating faces could represent an increased processing need (Olivares et al., 2015). As increased processing need is linked to negative aesthetic judgments (Musch & Klauer, 2003), such a process may explain the uncanniness of deviating faces. However, results on face-sensitive neural activity and the uncanny valley is mixed, with some studies finding decreased activity (Rosenthal-von der Pütten, Krämer, Maderwald, Brand, & Grabenhorst, 2019; Schindler, Zell, Botsch, & Kissler, 2017) while others find an increase (Kim et al., 2016). The exact association between increased face-related processing need and the uncanny valley is unclear.

Expectation violating and predictive coding

Predictive coding. The brain is in an efficient equilibrium when internal generative models and predictions of the world are in tune with sensory input, while a discrepancy between prediction and sensory information elicits a prediction error (Friston, 2010, Keller & Flogel, 2018). Prediction errors are operationalized as increased neural activity when sensory input conflicts with previously learnt patterns (Fiser et al., 2018; Makino & Komiyama, 2015; Meyer & Olson, 2011).

The N400 ERP component is a neural correlate of prediction errors (Kutas & Federmeier, 2011). N400 components are usually observed for unexpecting events or semantic errors in sentences (Kutas & Hillyard, 1980). N400 effects have also been observed for face stimuli, for example for mismatches between familiar faces and learnt context primes (Jemel, George, Olivares, Fiori, & Renault, 1999; Olivares & Iglesias, 2010; Olivares, Iglesias, & Maria, 1999; Wiese & Schweinberger, 2008) rooted in activity in face-sensitive areas (Olivares et al., 2018).

Prediction error and the uncanny valley. Prediction errors could occur when experience-driven expectations of human appearance and behaviour contradict observations of an imperfect artificial humanoid: Discrepancies between androids' humanlike appearances and mechanical motions elicit N400 components (Mustafa, Guthe, Tauscher, Goesele, & Magnor, 2017; Urgen, Kutas, & Saaygin, 2018). However, the research did not measure the stimuli's uncanniness, leaving its link to aversive emotional reactions unclear. No differences in N400 amplitudes between android and robot or human images (Urgen et al., 2018), despite images of androids typically being uncanny (Diel et al., 2022), furthermore muddles the association between N400 amplitudes and the uncanny valley.

N400 components as indicators of prediction errors are typically investigated in an experimental setup in which an unexpected stimulus follows a context stimulus which cues the "prediction", e.g., a semantically surprising ending in a sentence (Kutas & Federmeier, 2011), or an identity-mismatched face followed by a face identity cue (Olivares et al., 2018). However, as stimuli can appear uncanny without a preceding "prediction cue", a lack of N400 component effects would not support prediction error as an explanation of uncanniness.

In summary, N400 amplitudes as indicators of prediction error do not seem to fully explain the uncanny valley. Thus, it is yet unclear how well prediction error, operationalized as an increased N400 response, can predict uncanniness of uncanny still images.

This work aims to critically investigate both the specialised processing and prediction error theories in relation to the uncanny valley.

Biologically non-typical faces

Naturally occurring deviating faces, such as facial disfigurements or faces containing anomalies like scars, elicit higher activity in the amygdala and the FFA(Workman et al., 2021; Hartung et al., 2019), and are evaluated negatively similarly to uncanny faces (Diel &

MacDorman, 2021). Specialized processing or prediction error mechanisms may explain the negative evaluation of untypical biological faces. In that case, the perception of untypical biological faces would show behavioural and neural reactions analogous to those of uncanny faces. Hence, this work additionally investigates the inversion effect and neural correlates of untypical biological faces in the context of the theories discussed above.

Experiment 14

This work aims to investigate neural correlates of the uncanny valley in faces. Uncanny (mismatching and Thatcher) faces were compared to fully human faces and virtual faces. In addition, biologically non-typical faces were included as stimuli to investigate whether the observed effects apply to naturally occurring deviating faces and whether the theories investigated could also explain negative attitudes towards people with biologically non-typical faces.

Research question and hypotheses

For the first part of the study, behavioural and neural correlates of locally distorted faces (Thatcher and mismatch) are investigated. Behavioural data is used to replicate an uncanny valley in faces. The role of configural information to assess face uncanniness is investigated through face inversion. "Mismatch" faces are used (human faces with eyes and mouth swapped with those of unreal avatar faces) because feature realism mismatch is associated with uncanniness (MacDorman & Chattopadhyay, 2017). In addition, Thatcher faces are used given their use in the investigation of global inversion in the detection of configural deviation. The presence of an uncanny valley will be tested by investigating whether a cubic N-shaped function akin to Mori's (2012) proposed plot can best explain uncanniness data as a function of human likeness for upright, but not inverted faces:

1. Uncanniness of mismatch and Thatcher faces, but not of normal or unreal faces, is decreased when faces are presented inverted instead of upright.
2. A cubic function of human likeness can best explain uncanniness of upright faces, but not when faces are inverted.

Neural correlates of face-sensitive processing are first investigated by using N170 as a marker of configural processing, and P100 as a marker of featural processing. Specifically, if N170 amplitude is a marker of configural distortion, then it should show increased amplitudes for mismatch and Thatcher faces, albeit not when configuration is disrupted through inversion. Meanwhile, P100 as a marker of featural processing should be sensitive to facial distortion (mismatch and Thatcher faces) regardless of orientation:

3. N170 amplitude is increased for mismatch and Thatcher faces compared to normal or unreal faces, but only when faces are presented upright.
4. P100 amplitude is increased for mismatch faces, but not for Thatcher faces, compared to normal or unreal faces.

Neural correlates of predictive coding as a potential explanation of uncanniness are investigated by using N400 as an indicator of prediction errors. Specifically, it is investigated whether N400 amplitudes are sensitive to facial distortions which are also expected to be the most uncanny, i.e., mismatch and Thatcher faces when presented upright, but not inverted:

5. N400 amplitudes are increased for mismatch and Thatcher compared to normal or unreal faces, but not when faces are presented inverted.

In addition, more severely distorted faces (including changed sizes and positions of facial features) will be used as a separate stimulus set. Hence, the same hypotheses as above will also be tested in relation to naturally untypical and distorted faces.

Finally, indicators of both (featural and configural) face processing and prediction errors are used as predictors of face uncanniness across all face conditions:

6. Face uncanniness is best predicted by a) neural correlates of face-sensitive processing, and b) neural correlates of prediction errors.

Methods

Participants

Power analysis was conducted using Pangea®, and revealed that a set of 85 participants is sufficient given a medium effect size of $d = 0.5$ (Cohen, 1988). A total of 85 participants were recruited for the experiment. Forty-nine participants were recruited and tested at the Ruhr-University Bochum, and 36 at Cardiff University. Participants were aged $M_{age} = 21.72$ ($SD_{age} = 1.96$). Participants gave informed consent either electronically or on paper before the experiment began. The study was approved by the Ruhr University's Ethics Committee on 8th May 2019 (No. 548) and by the Cardiff University's Ethics Committee on February 1st 2021 (EC.21.01.12.6246G). All methods were performed in accordance with relevant guidelines and regulations.

Stimuli

"Deviating" faces are here defined as either artificially created or naturally occurring atypical, anomalous, or distorted faces with an intact global configural face pattern (e.g., scrambled faces or faces with swapped eye and mouth positions are not considered deviating).

For the first part of the study, sets of 50 (25 male, 25 female) normal, Thatcher, mismatch, and unreal faces were selected. Normal faces were cropped versions of faces from the Chicago face database (chicagofaces.org) and the Aberdeen face set from the Psychological Image Collection at Stirling (PICS, http://pics.stir.ac.uk), Thatcher faces were a different set

of faces from the same databases, with eyes and mouth locally inverted. Unreal faces were faces of virtual avatars generated using the MakeHuman® software (makehuman 1.1.1, http://www.makehumancommunity.org) which allows generating a large set of standardized artificial human faces, while also ensuring that the presented virtual faces were unknown to the participants. Finally, mismatch faces were created by placing eyes and mouth of unreal faces onto a new set of normal human faces taken from the previously mentioned databases.

Because distortions in mismatch and Thatcher faces were only local (at the eyes and mouth), a set of 16 more severely configurally distorted faces were created by distorting relative proportions of facial features: eye-to-eye distance was increased, eye size reduced, and mouths elongated. Finally, a set of 16 faces of individuals with biologically non-typical faces were selected from various sources on the internet (e.g., individuals with Down's Syndrome or elephantitis). Distorted and biologically untypical stimuli were analogous to the stimuli used in previous research (Diel & MacDorman, 2021).

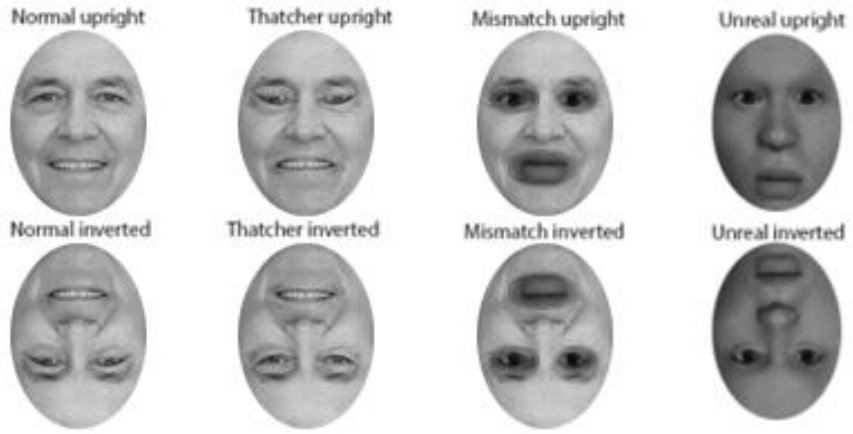
All images were cropped to remove clothing, hair, and ears in order to remove potential confounding variables unrelated to face processing. To minimize editing noise, images were greyscaled. Photoshop CS6 was used for editing faces.

As baseline stimuli for the identification of face-sensitive components in the EEG task, 100 images of houses were used, taken from the DalHouses database (Filliter, Glover, McMullen, Salmon, & Jognson, 2015). All stimuli were again cropped and greyscaled.

All stimuli were presented both upright and inverted. All face stimuli were unique for each face condition (except across upright and inverted conditions) in order to remove potential repetition effects on N170 amplitudes. Example stimuli are presented in *Figure 9.1*.

Figure 9.1

Example stimuli for each condition. Depicted human faces were artificially generated via StyleGAN (Karras et al., 2020) and were not used in the actual experiment. Experimental stimuli are not shown due to copyright reasons.

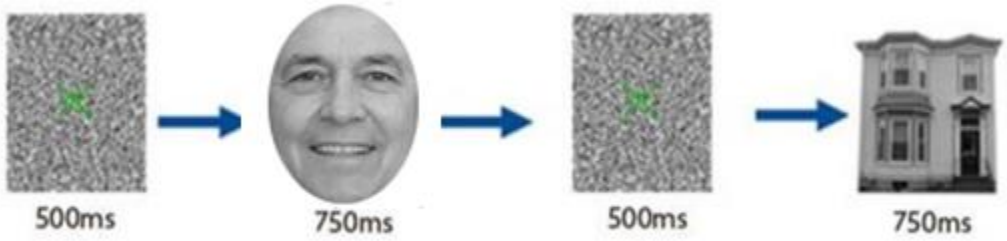


EEG Task

The EEG task consisted of 1200 trials, with three blocks containing 400 trials each. Each face stimulus appeared once per block, and stimuli were presented randomly. Each block took about 20 minutes, and participants were allowed to take breaks between the blocks. Break duration was decided by participants. A single trial consisted of a 500ms fixation cross on grey noise, followed by a 750ms face stimulus, followed again by a 500ms fixation cross and a 750ms house stimulus. Participants had to decide for each face whether it was upright or inverted to make sure participants were paying attention to the task. No participant had a correct response rate of below 70%, which would have been used as the threshold to exclude inattentive participants. An example trial is presented in *Figure 9.2*.

Figure 9.2

A single trial. Depicted human faces were artificially generated via StyleGAN (Karras et al., 2020) and were not used in the actual experiment. Experimental stimuli are not shown due to copyright reasons.



Rating Task

Participants rated each face stimulus of four scales: *eerie*, *creepy*, *disgusting*, and *humanlike*. According to a meta-analysis, the first three scales are commonly used scales in uncanny valley research, while the latter is most often used as a measure of perceived human likeness (Diel et al., 2022). Participants rated the stimuli on Likert scale ranging from 1 (fully disagree) to 7 (fully agree), by assessing how much they agreed with the statements *This face is eerie/creepy/disgusting/humanlike*. Participants could take their time responding to each statement.

EEG equipment and raw data processing

Sixty-four Ag-AgCL electrodes were arranged according to the standard international 10-20 system. In the German lab, a BrainAmp amplifier, and the BrainVision recording software were used to record the EEG signal (Brain Products GmbH, Gilching, Germany). In the UK lab, BioSemi ActiveTwo amplifiers and the ActiView recording software were used (BioSemi B.W., Amsterdam, Netherlands). FCz was used as a primary reference and impedances were kept below 10 kΩ. Participants wore a cap with electrodes attached on their scalps, and contact gel was applied on the electrodes.

BrainVision Analyzer 2.1 was used to process the EEG data (Brain Products GmbH, Gilching, Germany). All 200ms intervals with maximal value differences of 200 µV and a

minimum activity of 0.5 μV, or a low cut-off of 0.01 Hz and a high cut-off of 30 Hz were removed. Notch filters of 50 Hz were used. For all channels, Independent Component Analyses were conducted. Stimulus types were segmented into 800 milliseconds epochs (-200 to 600). After applying a baseline correction transformation (-200 to 0) and Current Source Density analysis, P100, N170, and N400 components were averaged for each stimulus type at relevant channels. Peak detection analysis was performed between 50 and 120 milliseconds (P100), 130 to 200 milliseconds (N170), and 300 to 500 milliseconds (N400). Range for N400 was wider given its wider observation interval in previous research (Kutas & Federmeier, 2011).

For each component, the electrodes T7, T8, TP7, TP8, TP9, TP10, PO7, and PO8 were selected, all found around parieto-temporal areas, and which have been associated with, structural face processing and especially then N170 (Eimer, 2011; Olivares et al., 2015). In addition, N400 amplitudes were measured at posterior parietal electrodes (P3, P4, Pz, POz, Oz) which have been associated with prediction errors in face processing and recognition (Olivares et al., 2015), and at frontal electrodes (F3, F4, F5, F6, F7, F8, Fz, AF3, AF4), which have been associated with prediction errors in relation to the uncanny valley (Mustafa et al., 2017; Urgen et al., 2018).

Procedure

Neurobehavioral Systems Presentation® was used to run both tasks (version 20.3, build 02.25.19). Each task began with a introduction and test trials. Participants could then continue with the EEG task if they had no further questions. Stimuli in *Figure 9.1* were used as test trial stimuli. The rating task was conducted after the EEG task to avoid face familiarity effects on the EEG data.

Data analysis and availability

RStudio® was used for data preparation and analysis. EEG and rating data were analysed separately. Linear mixed models and analyses of variance (ANOVAs) or their non-parametric counterparts have been used for the main analyses to ensure the generalisability of the results (Yarkoni, 2022) and to avoid the stimuli-as-a-fixed-effect fallacy that has been an issue in imaging research (Westfall, Nichols, & Yarkoni, 2016). For uncanniness analyses, linear mixed models were constructed with linear, quadratic, and cubic functions of human likeness as fixed factors and participants as random factors. For electrode analyses, ERP amplitudes have been analysed with linear mixed models with electrode and stimulus type and fixed effects and participants as random effects. For post-hoc tests, Bonferroni-adjusted p-values are reported. The packages lme4 and lmerTest have been used (Bates, Mächler, Bolker, & Walker, 2015). Data, stimuli, and analysis are available at https://osf.io/gv6ar.

Results

Uncanniness ratings

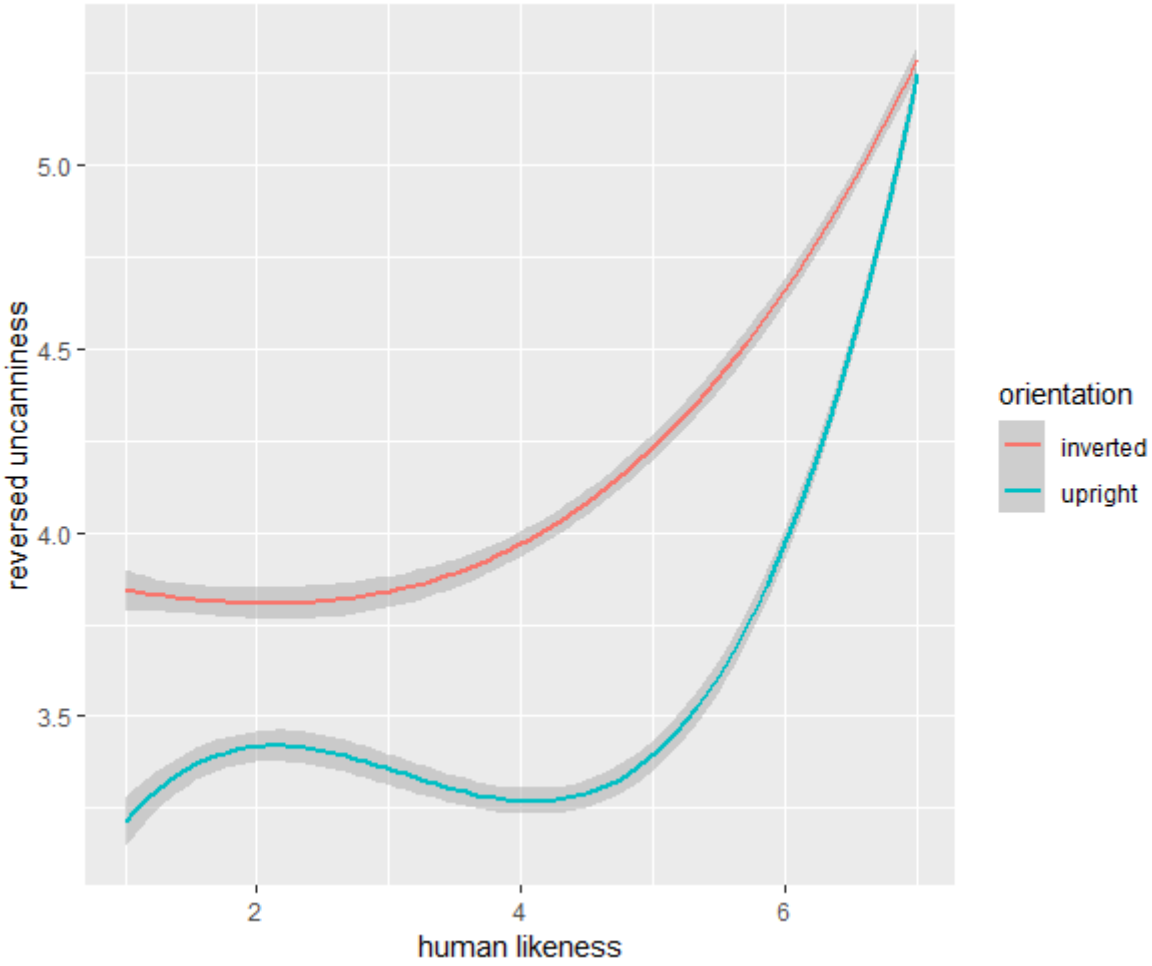
Rating scales. The scales *eerie*, *uncanny*, and *disgusting* were combined to an uncanniness index by calculating the mean across the three scales (α = .772).

Uncanny valley function. Results show that a quadratic function of human likeness ($t(304440) = -46.72$, $p < .001$, $R^2_{adj} = .466$) could explain uncanniness better than a linear function ($t(304440) = -78.92$, $p < .001$, $R^2_{adj} = .466$; $\chi^2 = 2108.1$, $p < .001$), and a cubic function ($t(30440) = -8.688$, $p < .001$, $R^2_{adj} = .498$) could explain the data better than a quadratic ($\chi^2 = 75.39$, $p < .001$) or linear function ($\chi^2 = 2183.5$, $p < .001$) for upright faces.

A cubic function was plotted with face orientation as an interaction variable. The interaction between cubic human likeness and face orientation was significant ($t(30410) = -12.953$, $p < .001$, $R^2_{adj} = .534$). The data is plotted in *Figure 9.3*.

Figure 9.3

Uncanniness plotted against human likeness for upright and inverted faces. Reversed uncanniness plotted against human likeness, divided by upright and inverted faces. Lines indicate best cubic fits, and grey areas show standard error ranges.

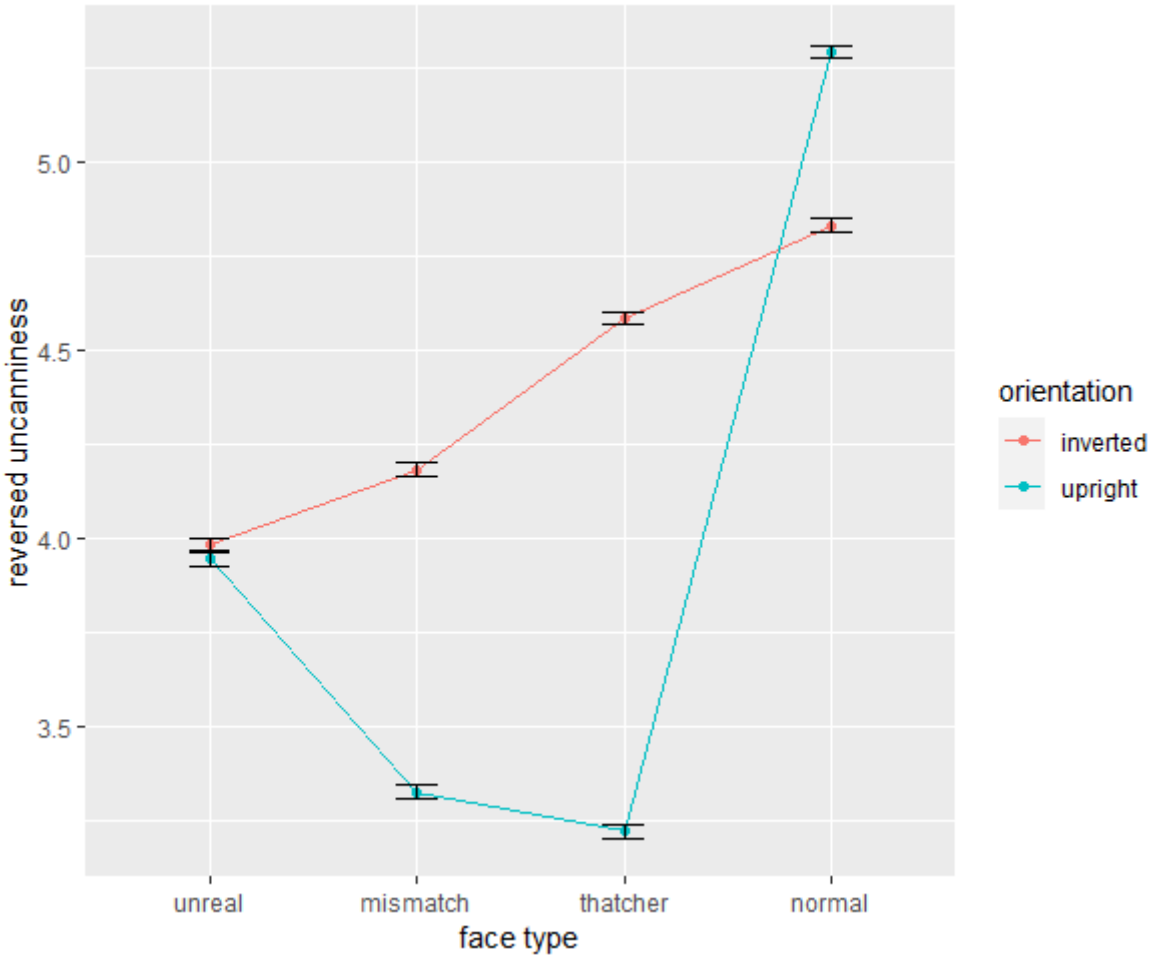


Uncanniness ratings across face conditions. Because the scale ordinal scales were used to measure uncanniness, the non-parametric Kruskal-Wallis test was used to investigate the interaction between face type and orientation on uncanniness ratings. Face condition significantly affected uncanniness ratings ($\chi^2 = 174.21$, $p < .001$). For post-hoc analysis, Wilcoxon rank sum tests with Bonferroni adjustments have been performed: Normal faces were significantly less uncanny than upright Thatcher ($W = 523.5$, $p_{adj} < .001$, $\delta = -.84$),

upright mismatch (W = 623, p_{adj} < .001, δ = -.*82*), and upright unreal faces (W = 1174.2, p_{adj} < .001, δ = -.*68*). In addition, while inverted normal faces were significantly more uncanny than upright normal faces (W = 2030.5, p_{adj} < .001, δ = -.*36*), inverted Thatcher faces were less uncanny than their upright counterparts (W = 5328.5, p_{adj} < .001, δ = *.52*), just like for mismatch faces (W = 4450.5, p_{adj} < .001, δ = *.33*). No difference was observed for upright and inverted unreal faces (W = 3349, p_{adj} = .499, δ = *.02*). The data is summarized in *Figure 9.4*.

Figure 9.4

Uncanniness ratings across face types and orientation. Average uncanniness ratings across face types and conditions. Error bars depict standard errors.

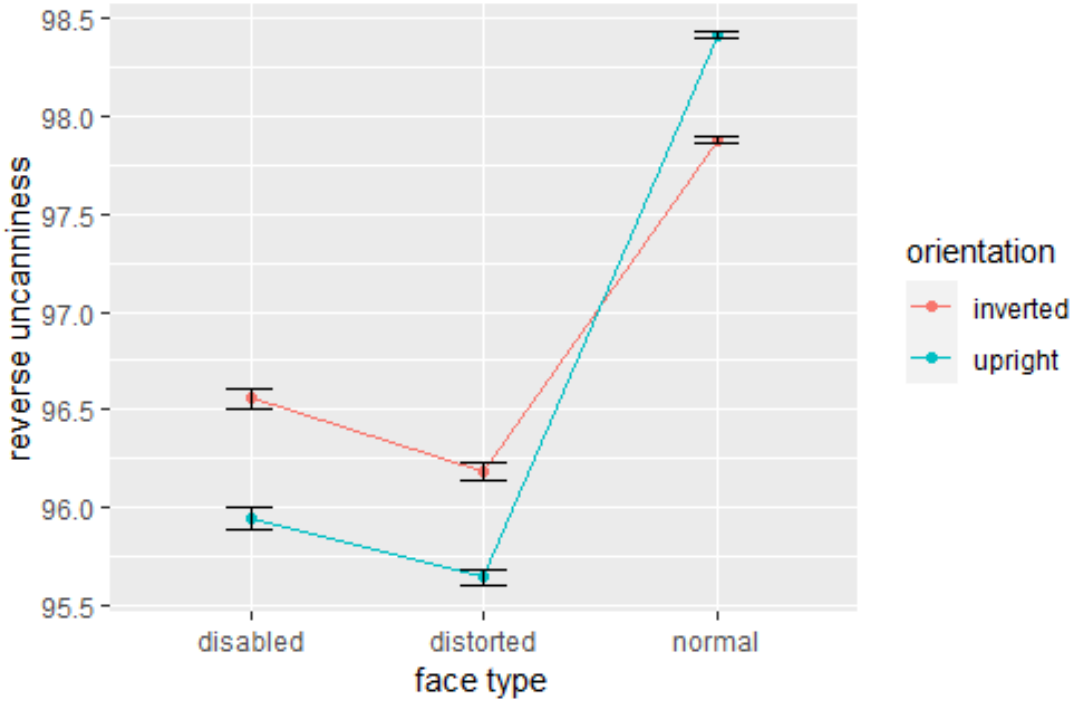


Uncanniness of distorted and biologically non-typical faces

Uncanniness of distorted and biologically non-typical faces (here referred to as "disabled) have been investigated analogous to the uncanniness of mismatched and Thatcher faces. Kruskal-Wallis tests showed significant differences between conditions (χ^2 = 174.83, p < .001). Post-hoc Wilcoxon tests revealed that normal upright faces were significantly less uncanny than disabled (W =197.5, p_{adj} < .001, δ = *.8*) or distorted upright faces (W = 226.5, p_{adj} < .001, δ = *.89*). While normal upright faces were less uncanny than normal inverted faces (W = 1215, p_{adj} < .001, δ = *.37),* disabled *(W = 2982.5*, p_{adj} = .004, δ = *.18) and distorted (W = 2857.5*, p_{adj} = .009, δ = *.18)* upright faces were less uncanny than their inverted counterparts. Data is summarized in *Figure 9.5.*

Figure 9.5

Uncanniness ratings across face type and orientation. Average uncanniness ratings across face types (disabled/biologically non-typical, distorted, normal) and conditions. Error bars depict standard errors.

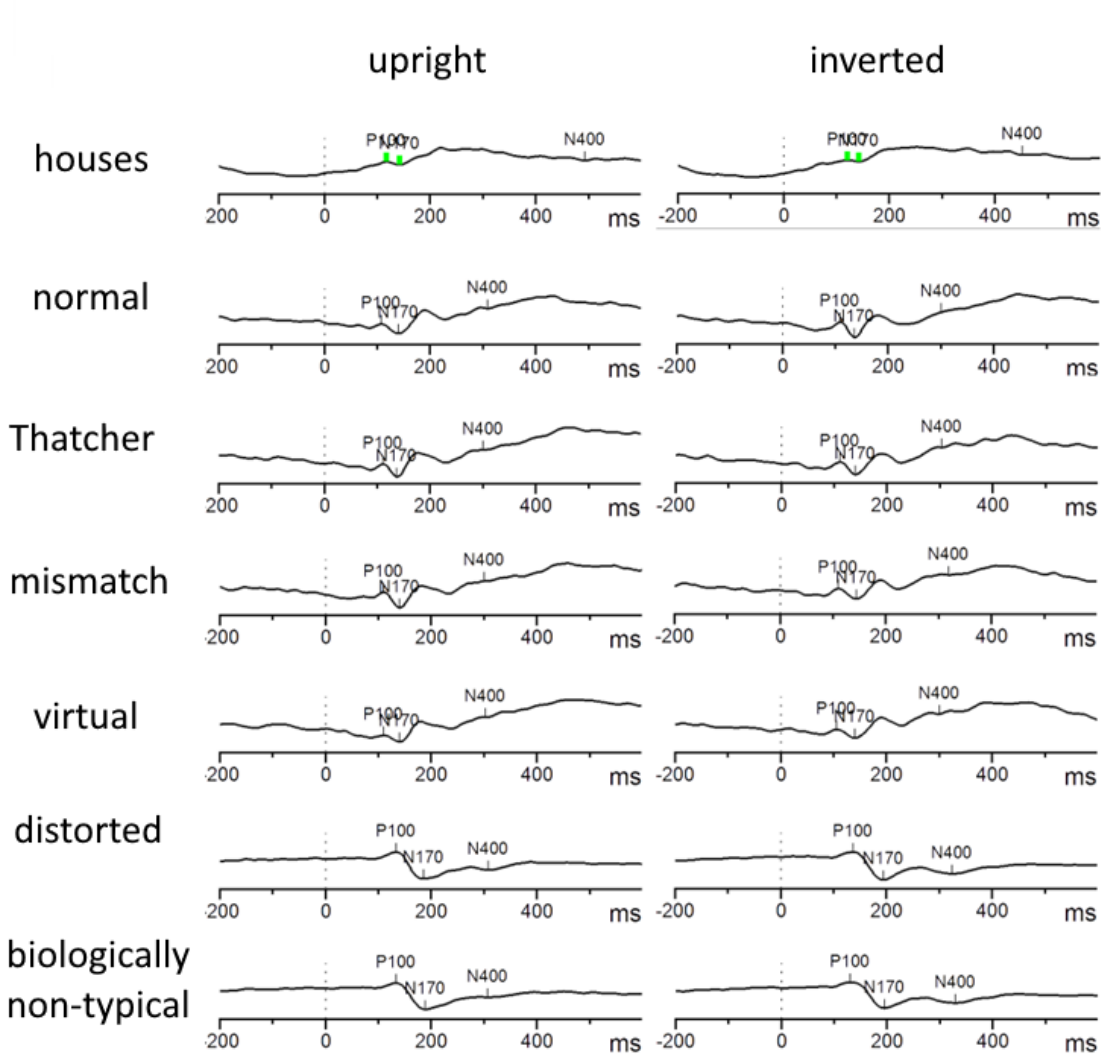


EEG analysis

Extracted P100, N170, and N400 amplitudes are visualized in *Figure 9.6* for the face-sensitive electrode TP8.

Figure 9.6

P100, N170, and N400 amplitudes across conditions for example electrode TP8.



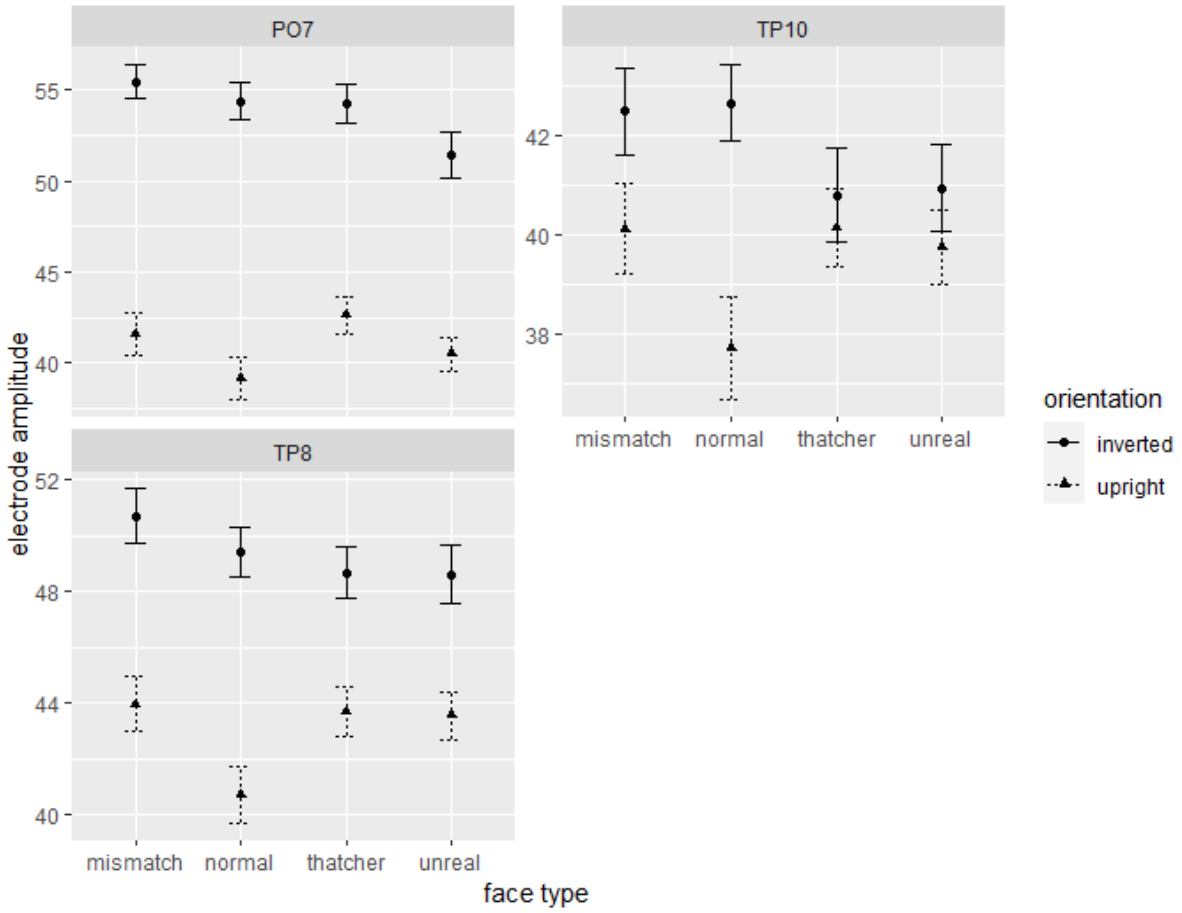
Note that while P100 and N170 amplitudes were visible at face-sensitive electrode sites across face conditions, N400 amplitudes were not visible at any sites for any condition. Hence, the experimental setup may not have been designed to capture N400 effects. N400 analyses should be interpreted with caution.

Face sensitivity. To validate face sensitivity of the ERPs of interest, amplitude of the P100, N170, and N400 between face and house stimuli have been compared for all relevant electrode sites through t-tests.

N170. ANOVA with face type, orientation, and electrode site as within-factors were conducted for the main analyses. Effects of face type and orientation on N170 amplitudes are reported across all electrode sites. Significant interactions between face and orientation and post-hoc comparisons are reported only at electrode sites with significant interactions. Sphericity-corrected statistics are presented when Mauchly's assumption for Sphericity was violated. The data for each electrode site (using reversed values) is summarized in *Figure 9.7.*

Figure 9.7

Reversed N170 amplitudes across conditions. Mean reversed N170 electrode amplitudes (microvolts) at electrode sites with significant interactions between face type and orientation. N170 amplitude values have been reversed to indicate that higher values correspond to stronger (negative) amplitudes. Error bars depict standard errors.



Significant interaction effects between face and orientation were revealed at electrode sites TP8 ($F_{(3,126)} = 4.19$, $p < .001$, $\eta^2 = .001$), TP10 ($F_{(3,126)} = 5.987$, $p < .001$, $\eta^2 = .002$), and PO7 ($F_{(3,126)} = 4.741$, $p = .001$, $\eta^2 = .001$).

TP8. Significantly increased N170 amplitudes at site TP8 for Thatcher ($t(294) = 2.344$, $p_{adj} = .02$, $d = 0.51$), mismatch ($t(294) = 2.105$, $p_{adj} = .018$, $d = 0.45$), and unreal faces ($t(294) = 1.975$, $p_{adj} = .049$, $d = 0.43$) compared to normal faces were observed. These difference were not present when faces were inverted (normal vs Thatcher: $t(294) = -0.548$, $p_{adj} = .88$; normal vs mismatch: $t(294) = -0.953$, $p_{adj} = .051$; normal vs unreal: $t(294) = -0.605$. $p_{adj} = .82$). Inverted faces elicited significantly higher amplitudes than their upright counterparts for normal ($t(294) = -6.352$, $p_{adj} < .001$, $d = 0.1.37$), Thatcher ($t(294) = -3.461$, $p_{adj} = .001$, $d = 0.75$), mismatch ($t(294) = -5.201$, $p_{adj} < .001$, $d = 1.12$), and unreal faces ($t(294) = 3.773$, $p_{adj} < .001$, $d = 0.43$, $d = 0.81$).

TP10. Amplitudes were significantly higher for Thatcher ($t(294) = 2.08$, $p_{adj} = .038$, $d = 0.45$) and mismatch ($t(294) = 1.677$, $p_{adj} = .047$, $d = 0.36$), albeit not unreal faces ($t(294) = 1.57$, $p_{adj} = .12$), compared to normal faces. These differences were not observed when faces were inverted (normal vs Thatcher: $t(294) = -1.471$, $p_{adj} = .21$; normal vs mismatch: $t(294) = -0.164$, $p_{adj} = 1$; normal vs unreal: $t(294) = -1.38$. $p_{adj} = .25$). Normal inverted faces elicited higher amplitudes than normal upright faces ($t(294) = -3.917$, $p_{adj} < .001$, $d = 0.84$); however, this difference between inverted and upright faces was not observed for any other face type (Thatcher: $t(294) = -0.366$, $p_{adj} = 1$; mismatch: : $t(294) = -2.076$, $p_{adj} = .08$; unreal: : $t(294) = -0.966$, $p_{adj} = .67$).

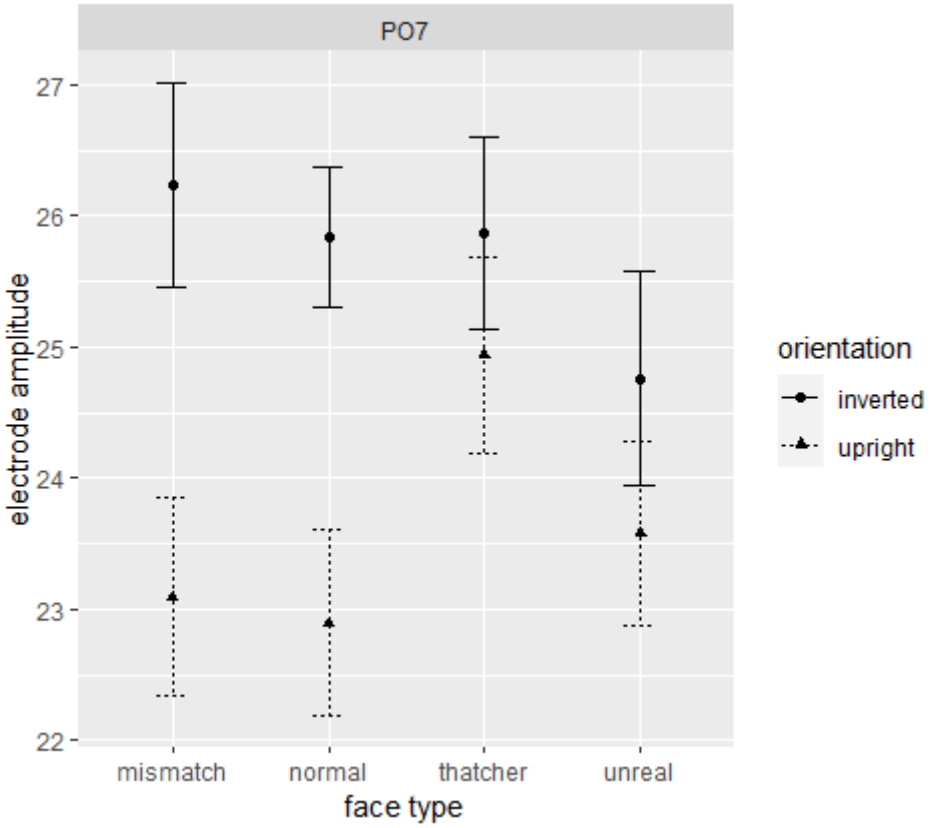
PO8. Significantly higher amplitudes for Thatcher upright compared to normal upright faces ($t(294) = 2.217$, $p_{adj} = .0274$, $d = 0.45$), but not for mismatch upright ($t(294) = 0.894$, $p_{adj} = .372$) or unreal upright faces ($t(294) = 1.517$, $p_{adj} = .065$) were observed. No differences were found for inverted faces (normal vs Thatcher: $t(294) = -0.153$, $p_{adj} = 1$; normal vs mismatch: $t(294) = -0.72$, $p_{adj} = 0.71$; normal vs unreal: $t(294) = -1.819$, $p_{adj} = .11$). Inverted faces elicited significantly higher amplitudes than their upright counterparts for normal ($t(294) = -9.602$, $p_{adj} < .001$, $d = 0.84$), Thatcher ($t(294) = -7.231$, $p_{adj} < .001$, $d = 0.08$), mismatch ($t(294) = -8.805$, $p_{adj} < .001$, $d = 0.81$), and unreal faces ($t(294) = -6.89$, $p_{adj} < .001$, $d = 0.21$).

P100. Data of P100 amplitudes for electrodes with significant main or interaction effects are summarized in *Figure 9.8*.

Figure 9.8

P100 amplitudes across conditions.

Mean P100 electrode amplitudes (microvolts) at electrode site with significant interactions between face type and orientation. Error bars depict standard errors.



Significant interactions were found between face type and orientation ($F_{(3,126)}$ = 2.698, p = .049, η^2 = .001) only at electrode site PO7. While upright Thatcher faces elicit a higher P100 amplitude than upright normal faces (t(306) = -1.981, p_{adj} = .048, d = 0.42), upright mismatch (t(306) = -0.19, p_{adj} = .425) or unreal faces (t(306) = -0.663, p_{adj} = .508) do not. No comparisons were significant when faces were inverted (normal vs Thatcher: t(306) = -0.023, p_{adj} = 1; normal vs mismatch: t(306) = 0.369, p_{adj} = 1; normal vs unreal: t(306) = 1.046, p_{adj} = .443). Inverted faces elicited stronger P100 amplitudes than upright faces for normal (t(306) = 2.834, p_{adj} = .01, d = 0.61) and mismatch faces (t(306) = 3.055, p_{adj} = .005, d = 0.64), but not for Thatcher (t(306) = 0.914, p_{adj} = .723) or unreal faces (t(306) = 1.146, p_{adj} = .505).

ERPs of distorted and biologically non-typical faces

N170**.** Interaction effects between face type and orientation were observed at T7 ($F_{(2,62)}$ = 5.946, p = .009, η^2 = .03), TP7 ($F_{(2,62)}$ = 3.171, p = .009, η^2 = .03), TP8 ($F_{(2,62)}$ = 5.946, p = .009, η^2 = .03), TP9 ($F_{(2,62)}$ = 4.085, p = .021, η^2 = .008), TP10 ($F_{(2,62)}$ = 9.66, p < .001, η^2 = .006), PO7 ($F_{(2,62)}$ = 7.559, p = .001, η^2 = .004), and PO8 ($F_{(2,62)}$ = 6.553, p = .003, η^2 = .002).

T7. Upright normal faces elicited weaker amplitudes than upright biologically non-typical faces (t(155) = -5.657, p_{adj} < .001, d = 1.41), but not upright distorted faces (t(155) = -1.188, p_{adj} = .473). Inverted normal faces did not elicit weaker amplitudes than biologically non-typical (t(155) = -0.878, p_{adj} = .763) or distorted (t(155) = -0.682, p_{adj} = .993) ones. Finally, upright biologically non-typical faces elicited stronger amplitudes than inverted counterparts (t(155) = 3.77, p_{adj} < .001, d = 0.94), which was not observed for normal (t(155) = -1.009, p_{adj} = .315) or distorted (t(155) = -0.502, p_{adj} = .616) faces.

TP7. Upright normal faces elicited weaker amplitudes than upright biologically non-typical (t(155) = -3.403, p_{adj} = .002, d = 0.85) and distorted faces (t(155) = -2.383, p_{adj} = .0367, d = 0.6) but not when inverted (biologically non-typical: t(155) = 0.323, p_{adj} = 1; distorted: t(155) = -0.828, p_{adj} = .818). Again, upright biologically non-typical faces elicited stronger amplitudes than inverted biologically non-typical faces (t(155) = 3.088, p_{adj} = .001, d = 0.77), which was not observed for distorted (t(155) = 0.917, p_{adj} = 1) or normal (t(155) = -0.638, p_{adj} = 0.524) faces.

TP8. Normal upright faces again elicited weaker amplitudes than biologically non-typical (t(155) = -2.621, p_{adj} = 0.019, d = 0.66) and distorted faces (t(155) = -2.446, p_{adj} = 0.031, d = 0.61), but not when inverted (biologically non-typical: t(155) = 1.077, p_{adj} = 1; distorted: t(155) = -0.637, p_{adj} = 1).

TP9. Normal upright faces elicited weaker amplitudes than biologically non-typical ($t(155) = -3.315$, $p_{adj} = .002$, $d = 0.83$), but not distorted faces ($t(155) = -1.69$, $p_{adj} = .186$), and not when inverted (biologically non-typical: $t(155) = 0.262$, $p_{adj} = .1$; distorted: $t(155) = -1.58$, $p_{adj} = .232$). Again, biologically non-typical upright faces elicited stronger amplitudes than inverted ones ($t(155) = 3.199$, $p_{adj} = .001$, $d = 0.8$), but not normal ($t(155) = -0.378$, $p_{adj} = .706$) or distorted faces ($t(155) = -0.268$, $p_{adj} = .789$).

TP10. Normal faces elicited weaker amplitudes than biologically non-typical faces when upright ($t(155) = -2.464$, $p_{adj} = .03$, $d = 0.62$), but not distorted faces ($t(155) = -1.935$, $p_{adj} = .11$), and not when inverted (biologically non-typical: $t(155) = 1.89$, $p_{adj} = 1$; distorted: $t(155) = 0.054$, $p_{adj} = 1$).

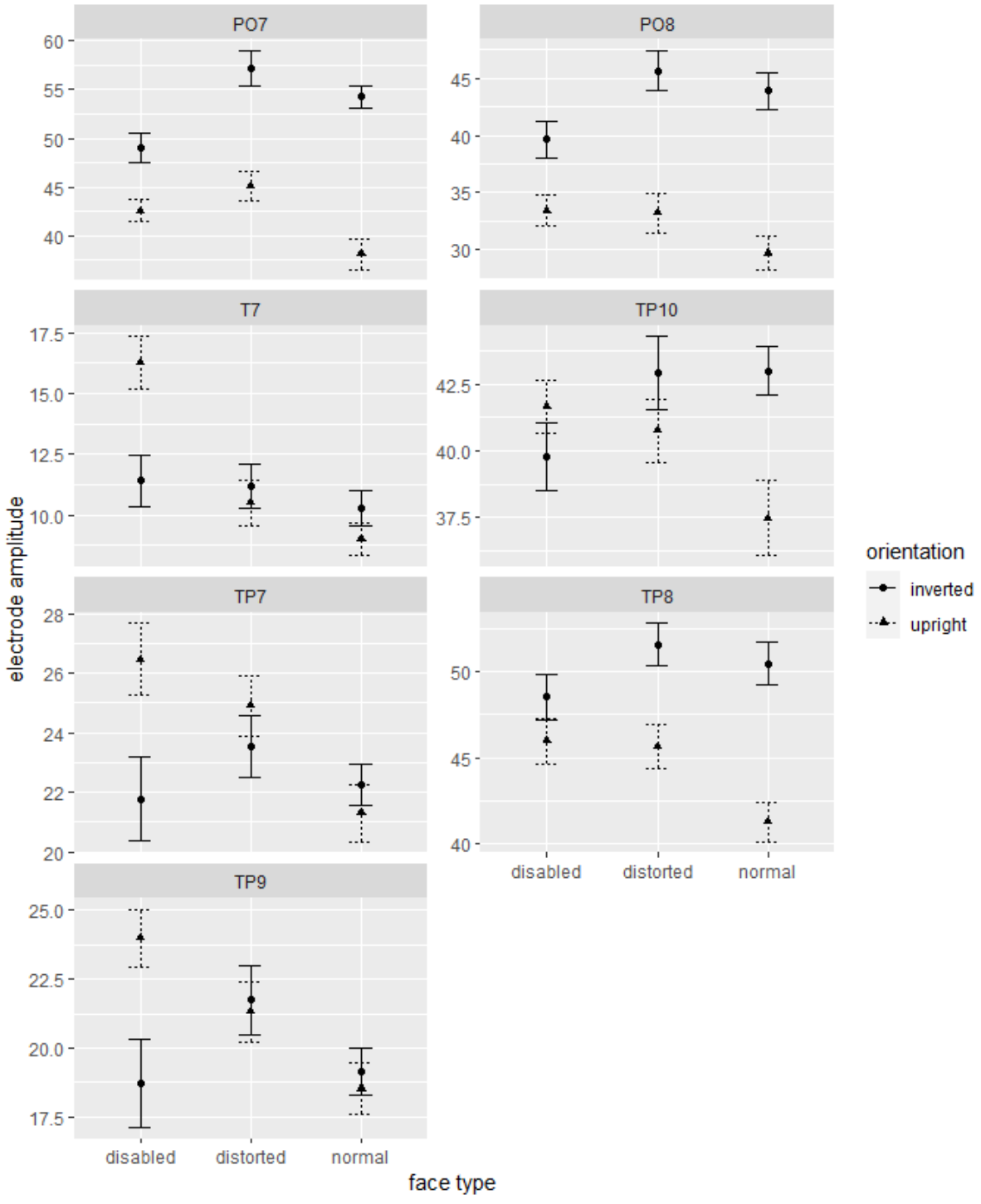
PO7. Normal upright faces did not elicit weaker amplitudes than biologically non-typical faces ($t(155) = -2.175$, $p_{adj} = .06$), but compared to distorted faces ($t(155) = -3.41$, $p_{adj} = .002$, $d = 0.85$), and not when inverted (biologically non-typical: ($t(155) = 2.515$, $p_{adj} = 1$; distorted: ($t(155) = -1.431$, $p_{adj} = .309$).

PO8. Normal upright faces did not elicit weaker amplitudes than biologically non-typical ($t(155) = -1.649$, $p_{adj} = .203$) or distorted ones ($t(155) = -1.557$, $p_{adj} = .243$), neither when inverted (biologically non-typical: $t(155) = 1.89$, $p_{adj} = 1$; distorted: $t(155) = 0.779$, $p_{adj} = 0.875$).

In summary, biologically non-typical (but not distorted) faces tended to elicit stronger N170 amplitudes compared to normal faces at multiple relevant electrode sites, albeit only when faces were presented upright. The data (using reversed N170 values) is summarized in *Figure 9.9*.

Figure 9.9

Reversed N170 amplitudes across conditions.Mean reversed N170 electrode amplitudes (microvolts) at electrode sites with significant interactions between face type and orientation. N170 amplitude values have been reversed to indicate that higher values correspond to stronger (negative) amplitudes. Error bars depict standard errors.



P100. Significant interaction effects were found at electrode site T7 ($F_{(2,62)} = 3.215$, $p = .046$, $\eta^2 = .018$), TP7 ($F_{(2,62)} = 3.14$, $p = .0498$, $\eta^2 = .012$), TP10 ($F_{(2,62)} = 4.826$, $p = .011$, $\eta^2 = .009$), and PO8 ($F_{(2,62)} = 6.12$, $p = .004$, $\eta^2 = .004$).

T7. Normal upright faces did not elicit weaker amplitudes compared to biologically non-typical ($t(178) = -0.791$, $p_{adj} = .215$) or distorted ones ($t(178) = 1.506$, $p_{adj} = .201$). When inverted, both biologically non-typical ($t(178) = 2.801$, $p_{adj} = .009$, $d = 0.64$) and distorted faces ($t(178) = 2.575$, $p_{adj} = .016$, $d = 0.6$) elicited stronger amplitudes than normal faces.

TP7. Normal upright faces did not elicit weaker amplitudes than biologically non-typical ($t(178) = 0.28$, $p_{adj} = .61$) or distorted faces ($t(178) = 1.247$, $p_{adj} = .321$), however biologically non-typical faces elicited stronger amplitudes than normal faces when inverted ($t(178) = 3.286$, $p_{adj} = .002$, $d = 0.45$). Distorted faces did not ($t(178) = 1.909$, $p_{adj} = .087$).

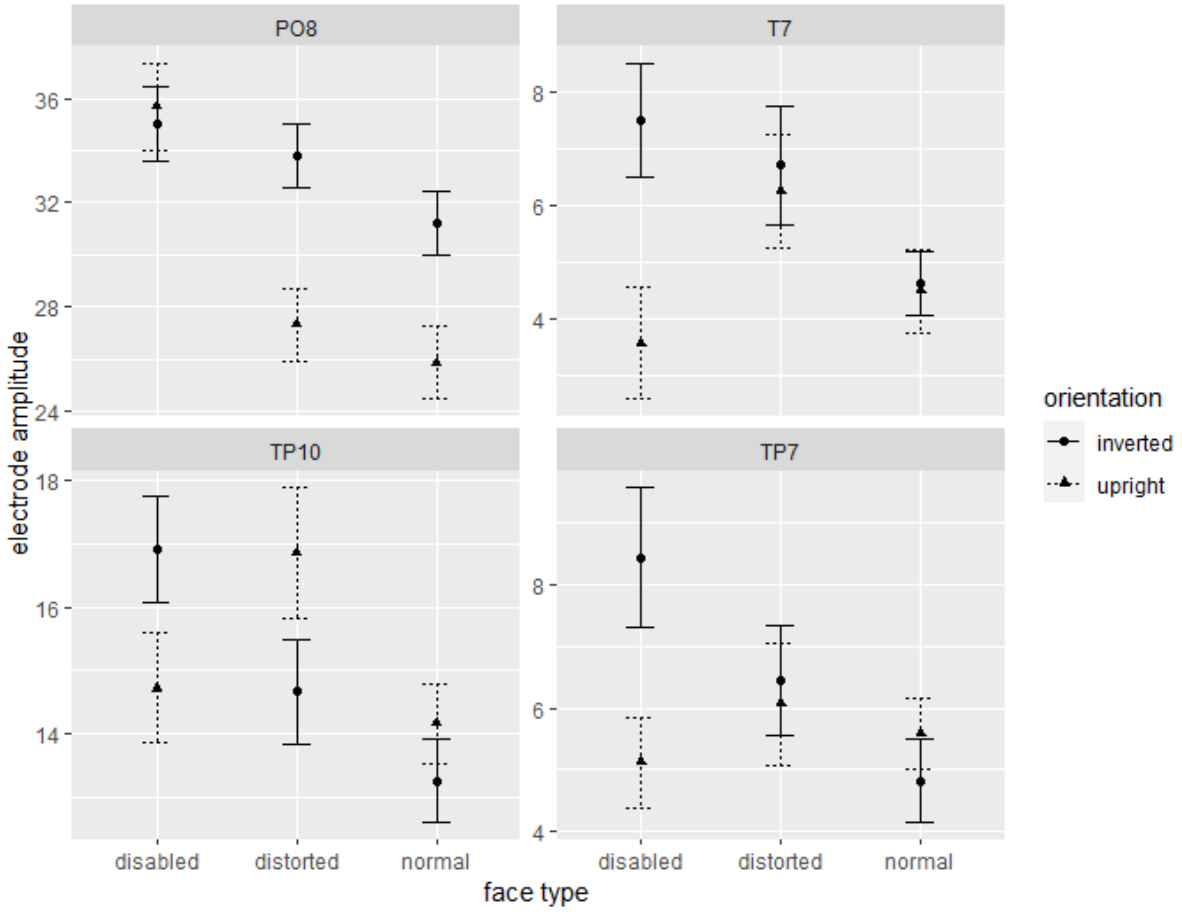
TP10. Upright biologically non-typical faces again did not elicit higher amplitudes than normal faces ($t(178) = 0.467$, $p_{adj} = .68$), upright distorted faces however did ($t(178) = 2.863$, $p_{adj} = .007$, $d = 0.66$). When inverted, biologically non-typical faces elicited higher amplitudes than normal faces ($t(178) = 2.86$, $p_{adj} = .007$, $d = 0.42$), distorted faces did not ($t(178) = 0.783$, $p_{adj} = .652$).

PO8. Upright biologically non-typical faces did elicit higher amplitudes than normal faces ($t(178) = 5.515$, $p_{adj} < .001$, $d = 1.28$), upright distorted faces however did not ($t(178) = 0.867$, $p_{adj} = .581$). The same was observed when faces were inverted (biologically non-typical: $t(178) = 2.637$, $p_{adj} = .014$, $d = 0.62$; distorted: $t(178) = 1.665$, $p_{adj} = .147$).

Data is summarized in *Figure 9.10*.

Figure 9.10

P100 amplitudes across conditions. Mean P100 electrode amplitudes (microvolts) at electrode sites with significant interactions between face type and orientation. Error bars depict standard errors.



N400 amplitudes

For both analyses, significant interaction effects between face type and orientation have been observed: Significant interaction effects were observed at sites TP8 ($F_{(3,126)} = 4.208$, $p = .01$, $\eta^2 = .002$) and TP10 ($F_{(3,126)} = 3.614$, $p < .021$, $\eta^2 = .003$).For the distorted and biologically non-typical face analysis, significant interaction effects were observed at TP7 ($F_{(2,62)} = 5.763$, $p = .006$, $\eta^2 = .02$), TP8 ($F_{(2,62)} = 3.838$, $p = .027$, $\eta^2 = .003$), TP9 ($F_{(2,62)} = 4.067$, $p = .021$, $\eta^2 = .01$), TP10 ($F_{(2,62)} = 3.639$, $p = .03$, $\eta^2 = .003$), PO7 ($F_{(2,62)} = 7.031$, $p = .004$, $\eta^2 = .006$), and PO8 ($F_{(2,62)} = 3.87$, $p = .037$, $\eta^2 = .002$). Note that N400 amplitudes were not visible in the graphical depictions across electrode sites and face conditions (e.g., see *Figure*

9.6). As the N400 is typically observed as a clear negative deflection[40], the lack of N400 amplitudes indicate that the experimental design was not suitable to elicit N400 effects. Interpretations of N400 analysis results should thus be treated with caution.

Neurophysiological predictors of uncanniness

To investigate the neurophysiological correlates of face uncanniness, a stepwise mixed model analysis was conducted with all ERP and electrode site combinations (38 in total) as fixed effect predictors of uncanniness, and participants as random effects. Analysis was done across all face conditions, including biologically non-typical and distorted faces. Stepwise analysis revealed that the combined four amplitudes (ERP component-electrode site) of N170-TP8 ($t(54) = -2.059$, $p = .044$), N170-PO8 ($t(63) = 3.64$, $p < .001$), P100-PO7 ($t(63) = -2.363$, $p = .021$), and N400-Fz ($t(96) = 3.102$, $p = .003$; all electrodes together: $R^2_{adj} = .22$).

Discussion

Results show that the uncanniness of upright faces could be best explained by a cubic function of human likeness indicative of an uncanny valley (Mori, 2012). However, a quadratic function best explained the relationship when faces were inverted. Inversion reduced the uncanniness of more uncanny face categories (mismatch, Thatcher, distorted, biologically non-typical), while it did not affect, or even increased the uncanniness of non-uncanny normal or unreal faces. These findings are in accordance with previous research of an "uncanniness inversion effect" (Chapters 2, 3, and 8). Given that inversion disrupts configural processing, the results can be explained by configural information used to asses face aesthetics (Diel & MacDorman, 2021; MacDorman et al., 2009) to the point that inversion can reduce or even eliminate the uncanny valley effect.

P100 amplitudes were inconsistently higher for deviating compared to normal faces when upright. P100 was also sensitive to differences between face conditions when faces were

inverted: Thus, P100 amplitudes seem sensitive to faced deviations even when configuration is disrupted due to inversion. This sensitivity may reflect the low-level processing of faces based on facial features preceding configural processing (Herrmann et al., 2004; Itier & Taylor, 2002). As P100 has shown to also account for configural information (Colombatto, & MacCarthy, 2017), the present results support that P100 responds to face processing with intact and disrupted configural processing, and that P100 amplitudes are increased for both configural and non-configural facial distortions or deviations.

N170 amplitudes were higher for mismatch, Thatcher, distorted, and biologically non-typical faces across various electrodes These differences were not present when faces were inverted. Thus, deviating faces elicited stronger N170 responses only when configural processing was undisturbed, indicating that higher relative amplitudes reflect the processing of configural deviations. Higher amplitudes for more distorted faces may reflect additional processing need for face configuration (Olivares et al., 2015).

Both behavioural and electrophysiological data support the conclusion that the uncanny valley is related to face-sensitive processing at least for uncanny face stimuli.

Configural information could be used to accurately assess face aesthetics (Chapter 2). Here, uncanniness of the most uncanny faces decreased while the uncanniness of the least uncanny (normal) faces increased with inversion. Specialized processing may develop from a need to differentiate stimuli on an individual level (Pascalis et al., 2011) and configural information can be important to detect subtle differences between stimuli (Chapters 2 to 4). A higher experience-dependent sensitivity for configural information could then increase the sensitivity to deviations from typical configural patterns, resulting in increased processing need and the sensation of eeriness or uncanniness for stimuli which are close to, but still deviating from, typical human appearance.

N400 amplitudes were measured according to previous research on the uncanny valley claiming to have found an association between the effect and the component (Urgen et al., 2018). However, N400 amplitudes were not visibly observed for any condition here. The experimental setup may not have been suitable to detect N400 effects. While interaction effects between face type and orientation were found, they were only present at face-sensitive electrode sites. The lack of visible N400 amplitudes cautions against the interpretations of the results. In any case, the results indicate that the experimental setup did not elicit N400 components and thus no meaningful N400 effects were observed. Because differences were only found at face-sensitive electrode sites, they may have resulted from preceding face-related neural processes.

Violation of expectation as prediction error is a common explanation in uncanny valley research (Kätsyri et al., 2015; MacDorman & Chattopadhyay, 2017; MacDorman et al., 2009; Mustafa et al., 2017; Saygin et al., 2012; Urgen et al. 2018; Wang et al., 2015). Prediction errors operationalized as N400 amplitudes have been previously associated with the uncanny valley in moving stimuli (Mustafa et al., 2017; Urgen et al., 2018). However, said research did not find N400 effects for still images of androids, despite androids being perceived as uncanny regardless of movement. N400 effects are found when observing expectation violations in the context of human (or humanlike) action and the understanding thereof (Amoruso et al., 2013; Bach, Gunter, Knoblich, Prinz, & Friederici, 2009; Proverbio & Riva, 2009; Shibata, Gyoba, & Suzuki, 2009; Urgen et al., 2018). Thus, the increased N400 amplitude observed in previous uncanny valley research (Urgen et al., 2018) may reflect an increased processing need for the interpretation of human action specifically rather than an effect underlying the uncanny valley in general.

N400 prediction error amplitudes are typically locally unspecific and observed across central parietal and frontal sites (Kutas & Federmeier, 2011), including in research on the uncanny

valley but only when using moving stimuli (Mustafa et al., 2017; Urgen et al., 2018). Thus, if the N400 component is used as an indicator of prediction error, the current results do not support prediction error as an explanation of the uncanny valley when applied to still images. As still images can fall into the uncanny valley, the results question the validity of previous research on the N400 as an indicator of prediction error in the uncanny valley.

Given the association of the N400 amplitude with semantic processing[40], the component may be unsuitable to study prediction error for still uncanny images. Alternative electrophysiological measures like rhythms may be suitable to study expectation violations in the uncanny valley: For example, gamma and theta/alpha oscillations have been associated with unpredictable stimuli in perceptual processing (Bastos et al., 2020; Michalareas et al., 2016; Uran et al., 2022). Thus, research linking prediction error to uncanny valley may focus on such rhythmic activity in the future. Furthermore, it is possible that the increased P100 and N170 components observed in this study reflect error signals caused by deviating stimuli. In sequence-based tasks, increased N170 amplitudes have been linked to unpredictable stimuli: for example, increased N170 amplitudes are observed for unpredicted face identities (Johnston et al., 2016). As predictive coding and processing fluency are not mutually exclusive and have been linked in the past (Robinson et al., 2018), increased amplitudes observed here may reflect error signal caused by a discrepancy between face schemata and deviating faces as a type of processing disfluency. In any case, the present results question the validity of previous research using N400 as an indicator of expectation violation in the uncanny valley.

Finally, significant neural predictors of face uncanniness were mostly spread around face-sensitive areas and components: N170 amplitudes at TP8 and PO8, and P100 amplitudes at PO7 best predicted uncanniness, in addition to N400 amplitudes at Fz. Thus, both early (P100) and mid-stage (N170) face-selective processing, as well as later, not face-selective

processing (N400) were relevant in predicting uncanniness. However, because N400 amplitudes were not observable in this study, the role of Fz N400 responses should be interpreted with caution.

The previous chapters found substantial evidence that uncanniness is caused by deviating stimuli, which is enhanced by specialization to the stimulus category and is associated with higher neural activity sensitive to these specialized categories. One other source of understanding the underlying mechanisms of a phenomenon like uncanniness effects is through investigating individual differences as predictors of the effect: For example, if an individual difference variable can predict uncanniness effects across categories, then uncanniness effects may have domain-general processing mechanisms. Using the stimuli of previous chapters, Chapter 10 will present such research.

Chapter 10: Individual differences in the uncanny valley: How deviancy aversion and disgust sensitivity relate to uncanny androids, strange places, and creepy clowns

Methods, experiment, and large portions of the introduction and discussion in this chapter is currently in review in the Journal *Computers in Human Behavior: Artificial Humans*.

Introduction

Investigating the effects of individual differences on the uncanny valley allows inferences on its cognitive mechanisms: For example, MacDorman and Entezari (2015) found that differences in disgust sensitivity predicted sensitivities to the uncanny valley, which would be expected by theories linking the uncanny valley to evolutionary disease avoidance mechanisms (MacDorman & Ishiguro, 2006). Although research on individual differences can provide important insights into the uncanny valley's underlying mechanisms, such research remains sparse (Abubshait, Momen, & Wiese, 2017; Lischetzke, Izydorczyk, Hüller, & Appel, 2017; MacDorman & Entezari, 2015; Sasaki, Ihaya, & Yamada, 2017). The aim of this work is to extend research on individual differences on the uncanny valley by focusing on previously ignored yet theoretically relevant personality variables (e.g., deviancy aversion; Gollwitzer et al., 2017), while also accounting for the uncanniness effects observed across various stimulus categories that have been found throughout this dissertation work. In the following, recent research on the uncanny valley (e.g., in inanimate categories) are discussed. Then, individual difference variables and their theoretical connections to the uncanny valley are described and research questions formulated.

Disgust sensitivity and disease avoidance

It has been suggested that the uncanny valley emerges due to evolved mechanisms of disease avoidance (MacDorman & Ishiguro, 2006). Evolved mechanisms of disease avoidance may drive contemporary negative attitudes towards people with disabilities due to a sensitivity to anomalous organic features (Park, Faulkner, & Schaller, 2003). As the function of disgust is

to avoid contamination (Rozin & Fallon, 1987), mechanisms of disease avoidance should be associated with disgust responses. Uncanny stimuli can indeed elicit disgust responses (Ho, MacDorman, & Pramono, 2008), and disgust sensitivity – a personality variable describing the relative strength of disgust reactions – is positively associated with the uncanny valley (MacDorman & Entezari, 2015). Furthermore, faces with disfigured features and pathological voices elicit negative responses akin to the uncanny valley (Chapter 7; Diel & MacDorman, 2021), supporting a connection between disease avoidance mechanisms and the uncanny valley.

Disease avoidance mechanisms may have evolved via a sensitivity and aversion to deviating organic features while having no relation to uncanniness in inanimate categories that possess no threat of contamination. Thus, disgust sensitivity should be associated with the uncanniness in deviating features in organic stimuli (e.g., faces, bodies, voices), while not being associated with uncanny deviations in inorganic stimuli (e.g., places, written text).

The Disgust Scale-Revised is a reliable measure of individual proneness to experience disgust reactions (Haidt, McCauley, & Rozin, 1994; Olatunji et al., 2007) and its Animal Reminder subfactor has been associated with the uncanny valley in past research (MacDorman & Entezari, 2015). Thus, the Disgust Scale-Revised and especially its Animal Reminder subfactor is a suitable candidate to measure disgust sensitivity. The questionnaire contains statements such as “It bothers me to hear someone clear a throat full of mucous” or “You see a man with his intestines exposed after an accident”. Statements are either rated on a “fully agree – fully disagree” scale or on a “not disgusting at all – extremely disgusting” scale.

Deviancy aversion

Deviations in simple patterns, like a sequence of geometric shapes tend to be devalued (Gollwitzer et al., 2017). As aversion to deviancy in simple patterns is associated to negative

attitudes towards individuals in statistical minorities or social deviancy, deviancy aversion is thought to be a domain-general mechanism (Gollwitzer et al., 2017; Gollwitzer et al., 2022). Negative evaluation of pattern deviancy may be caused by increased processing disfluency for statistically abnormal stimuli compared to a category prototype (Winkielman et al., 2003), or by violations in expectations in predictive coding (Friston, 2010). Meanwhile, the uncanny valley has been related to deviations in familiar categories driven by a higher sensitivity to anomalies due to specialized processing (Chapter 2 to 6; MacDorman & Chattopadhyay, 2016; Matsuda et al., 2012). Perceptual specialization may sensitize the processing of deviating information through a narrower range of acceptable variation, increasing effects of deviancy aversion in stimulus categories that are processed in a specialized manner, like faces (Kanwisher, 2000). As deviancy aversion is domain-general, uncanniness effects driven by deviancy aversion should occur independent of stimulus categories, encompassing animate or organic and inanimate or inorganic categories. Pattern deviancy is measured by showing disrupted or non-disrupted geometrical patterns which are rated on 9-scale “happy – unhappy”, “comfortable – uncomfortable”, and “content – discontent” scales following a “the above image makes me feel…” statement.

Pattern deviancy aversion has been measured by showing participants pairs of disrupted or non-disrupted patterns, and by asking participants of their levels of discomfort when viewing the images (Gollwitzer et al., 2017). As pattern deviancy aversion has been associated with devaluation of statistical minorities, social deviation, and people with disabilities (Gollwitzer et al., 2017; Gollwitzer et al., 2020). Thus, the pattern deviancy aversion measure is suitable to measure domain-general deviancy aversion that may underlie the uncanny valley.

Need for structure

Multiple theories on the uncanny valley imply that the phenomenon is related to violations of experience-based cognitive structures (see Lischetzke et al., 2017). Analogous to deviancy

aversion, individual differences exist in the degree at which individuals need to create unambiguous cognitive structures of the world (Neuberg & Newsom, 1993). Lischetzke et al. (2017) found that personal need for structure as a personality variable was associated with a higher sensitivity to the uncanny valley. Individual difference in the tolerance of disrupted cognitive structure may predict people's sensitivity to eeriness of artificial entities that disrupt expectations of human appearance and behaviour.

Although need of structure has been associated with humanlike entities specifically in the past (Lischetzke et al., 2017), the effect is supposed to be domain-general and is thus expected to predict uncanniness effects across animate and inanimate object categories.

Neuberg and Newsom (1993) developed and validated a personal need for structure questionnaire. Individuals with a high need for structure prefer to cognitively structure information in simple patterns, including the use of social stereotypes (Neuberg & Newsom, 1993), and have stronger negative responses to schema-inconsistent information (McGregor, Haji, & Kang, 2008). Need for structure questionnaire scores have been associated with the uncanny valley in the past (Lischetzke et al., 2017), and is thus a viable measure for the current study. The need for structure questionnaire uses "fully disagree – fully agree" scales on statements like "I enjoy having a clear and structured mode of life" or "I don't like situations that are uncertain."

Neuroticism (anxiety facet)

Uncanniness has been associated with fear and anxiety responses in past research (Ho et al., 2008). Neuroticism, a factor of the big five personality model, is associated with emotional instability, including sensitivity to anxiety and disgust responses and reaction towards threatening stimuli (Costa & McCrae, 1992; Digman, 1990). The anxiety facet of neuroticism

specifically has been found to sensitize uncanny valley reactions (MacDorman & Entezari, 2015).

Neuroticism is slightly associated with deviancy aversion (Golllwitzer et al., 2017) and could thus sensitize effects of deviancy aversion on uncanniness ratings of distorted stimuli. Furthermore, effects of neuroticism are expected to be independent of stimulus category.

The anxiety facet of neuroticism (here called anxiety-neuroticism) can be measured using the freely available International Personality Item Pool (Goldberg, 1999). It has been used to associate anxiety-neuroticism with the uncanny valley in the past (MacDorman & Entezari, 2015), and is thus a viable measure to replicate previous findings and extend them onto uncanniness effects beyond human(-like) appearance. The Anxiety-Neuroticism questionnaire uses “fully disagree – fully agree” scales on statements like “I worry about things” and “I get stressed out easily”.

Coulrophobia

Coulrophobia is considered a clinically significant fear of clowns (van Venrooji & Barnhoorn, 2017). On a subclinical level, 17.2% report at least being slightly afraid of clowns (vs 3.1% reporting being very afraid; Rapoport & Berta, 2019). Fear of clowns is a cross-cultural phenomenon affecting both children and adults (Meiri et al., 2017; Tyson et al., 2022), yet its aetiological mechanisms remain not well understood. Although some researchers suggested that fear of clowns may be caused by clowns falling into an uncanny valley due to their distorted humanlike appearance (Moore, 2012; Wang et al., 2015), yet without empirical investigation. Only recently research associated coulrophobia with the uncanny valley: Specifically, if the uncanny valley is understood as caused by entities deviating from typical human appearance, clowns’ exaggerated face and body proportions caused by make-up and costumes, then participants reported being distressed by clowns

because they depict such distorted human appearance (Tyson et al., 2023). However, Tyson et al. (2023) only asked participants on why they thought they were afraid of clowns; ratings of clown uncanniness and human likeness were absent. Hence, it remains unclear whether clowns actually fall into an uncanny valley, and whether this effect is especially pronounced in individuals with stronger self-reported fear of clowns.

If coulrophobia is associated with the uncanny valley, then people reporting a high fear of clowns should exhibit an "uncanny valley of clowns": Specifically, it is expected that for those individuals, clowns would tend to fall into the low point of an uncanny valley like function, akin to uncanny androids. In addition, clowns would be more uncanny and less humanlike compared to typical human stimuli.

Finally, an exploratory investigation into coulrophobia and the uncanny valley is presented: First, it is investigated whether coulrophobia can also predict the uncanniness of uncanny androids and distorted stimuli across categories. Second, it is investigated whether the effect of coulrophobia on clown uncanniness is mediated by further personality variables, such as deviancy aversion and disgust sensitivity.

Few measures of coulrophobia exist in the literature. A recent study developed and validated a Fear of Clowns Questionnaire (FCQ) by adapting it from a fear of spider questionnaire (Tyson et al., 2023). As the FCQ was capable of reliably measure individual differences in subclinical coulrophobia, it is a viable candidate for investigating associations between differences in fear of clowns and the uncanny valley. The FCQ uses "fully disagree – fully agree" scales on statements like "I would do anything to try to avoid a clown" and "If I saw a clown, I would feel very panicky".

In summary, the goal of the current study is to investigate the effect of individual differences on the uncanny valley and uncanniness effects beyond human(-like) stimuli. In addition, it is investigated whether and how coulrophobia is associated with the uncanny valley.

In addition to human(-like) face stimuli, including mechanical robots and androids, images of clowns are used. Furthermore, to investigate uncanniness effects in other categories, stimuli form previous studies investigated uncanniness effects are taken. Specifically, incrementally distorted bodies are taken from Chapter 8, normal and distorted places are taken from Chapter 6, typical, distorted, and pathological voices are taken from Chapter 7, and normal and orthographically distorted written sentences in familiar and unfamiliar languages are taken from Chapter 5.

Experiment 15

Research question and hypotheses

First, it was tested whether the data confirm expected uncanny valley and uncanniness effects, independently of individual differences. These validations co-function as replications of previously found uncanniness effects. The uncanniness validation hypotheses are as follows:

1. Across face stimuli, a polynomial (quadratic or cubic) function of human likeness can explain uncanniness better than a linear function (uncanny valley hypothesis)
2. Incremental face distortion increases uncanniness across face types, and more so for realistic (real and cartoon) compared to less realistic (drawing, CG, robot) face types (face uncanniness hypothesis)
3. Incremental body distortion increases body uncanniness (body uncanniness hypothesis)

4. Place distortion increases place uncanniness, and naturally deviating places are more uncanny than non-deviating places (place uncanniness hypothesis)
5. Distorted and pathological voices are more uncanny than typical voices (voice uncanniness hypothesis)
6. Distorted written text is more uncanny than non-distorted text, and this effect is more pronounced in more familiar languages (written text hypothesis)

Based on the elaborations above, hypotheses are formulated for each individual difference variable investigated:

For coulrophobia, it is expected that some clown stimuli fall into an uncanny valley (more uncanny and less humanlike compared to human stimuli), and that this effect is associated with coulrophobia disposition. In addition, exploratory analyses on coulrophobia will be performed.

1. Clown stimuli are more uncanny compared to typical human stimuli
2. Coulrophobia disposition predicts a polynomial function of human likeness and uncanniness for clown stimuli
3. Coulrophobia predicts uncanniness ratings for distorted face, body, and voice stimuli, but not place and text stimuli

Disgust sensitivity is expected to increase sensitivity to the uncanny valley (measured via a polynomial function of human likeness and uncanniness), and to increase uncanniness ratings for distorted organic (face, body, voice) stimuli, but not inorganic ones (place, text).

1. Disgust sensitivity predicts a polynomial function of human likeness and uncanniness across face stimuli
2. Disgust sensitivity predicts uncanniness ratings for distorted face, body, and voice stimuli, but not place and text stimuli

Deviancy aversion is expected to increase sensitivity to the uncanny valley and to increase uncanniness ratings of distorted stimuli across all investigated stimulus categories.

1. Deviancy aversion predicts a polynomial function of human likeness and uncanniness across face stimuli
2. Deviancy aversion predicts uncanniness ratings for distorted face, body, voice, place, and text stimuli

Need for structure is expected to increase sensitivity to the uncanny valley and to increase uncanniness ratings of distorted stimuli across all investigated stimulus categories.

1. Need for structure predicts a polynomial function of human likeness and uncanniness across face stimuli
2. Need for structure predicts uncanniness ratings for distorted face, body, voice, place, and text stimuli

Anxiety-Neuroticism is expected to increase sensitivity to the uncanny valley and to increase uncanniness ratings of distorted stimuli across all investigated stimulus categories.

1. Anxiety-Neuroticism predicts a polynomial function of human likeness and uncanniness across face stimuli
2. Anxiety-Neuroticism predicts uncanniness ratings for distorted face, body, voice, place, and text stimuli

Methods

Participants

As no previous studies investigated individual differences in various uncanniness effects, a large sample size was targeted. A total of 124 participants were recruited for this study. Participants were Cardiff University Psychology undergraduate students and US and UK

participants recruited via Prolific. Participants (M_{age} = 28.31, SD_{age} = 6.81) were 60 female, 56 male, five “other”, one preferred not to say, and two gender recordings were lost.

Materials

Questionnaires. Questionnaires used were the deviancy aversion measure (Gollwitzer et al., 2017), DS-R (Haidt et al., 1994; Olatunji et al., 2007), need for structure questionnaire (Neuberg and Newsom, 1993), anxiety facet of neuroticism (Goldberg, 1999), and the FCQ (Tyson et al., 2022). The deviancy aversion measure was directly adapted from the version provided by Gollwitzer et al., 2017). For the other questionnaires, scales from 0 to 100 on a “fully disagree – fully agree” scale for the questionnaire statements were used. Questionnaire statements were combined and presented in a randomized order.

Face stimuli. To create a range of stimuli varying across a scale of human likeness, seven different types of face stimuli in different levels of realism were used: typical human faces, cartoonized versions of real human faces, drawing-style rendered versions of real human faces, computer-generated faces, and robot faces, android faces, and clown faces.

To create ranges of face distortion across face types, human, cartoon, drawing, CG, and robot faces were incrementally distorted in three steps: no distortion, moderate distortion, high distortion. Distortions were created by increasing distance between the eyes and moving the mouth down.

Human stimuli were selected from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015). Cartoon and drawing faces were created by rendering different sets of human stimuli from the same database via cartoon character and sketch character tools of VanceAI toongineer (https://vanceai.com/toongineer-cartoonizer). CG faces were created via FACSGen. Robot and android faces were selected from previous research based on their human likeness and likability ratings (Mathur & Reichling, 2016). To have a set of uncanny

androids, least likable android stimuli specifically were selected. Finally, clown stimuli were selected using google image search. All clown images were selected to show frontal or side views of the face, and deliberately creepy clown designs were excluded to avoid confounding effects (deliberately creepy clown faces may rely on design choices that are not present in regular clowns, e.g., the use of blood or other intrinsically aversive stimuli). In total, 2 base real, cartoon, drawing, CG, and robot faces were used respectively, and included three levels of distortions (6 stimuli per real, cartoon, drawing, CG, and robot type). In addition, 16 images of androids and clowns were used, creating a total of 62 stimuli.

Body stimuli. Body stimuli were selected from the BEAST database (de Gelder & Van den Stock, 2011). Bodies were incrementally distorted be either extending or shrinking the length of arms and legs. Six body images were used with three levels of distortion each.

Place stimuli. Stimuli of physical places were taken from Chapter 6 on an uncanny valley of physical places. The set consisted of ten real places, five non-uncanny places and five of the most uncanny places. In addition, four pairs of virtual places were created using Roomstyler®, with one stimulus per pair being distorted and the other undistorted.

Voice stimuli. Voice stimuli were taken from Chapter 7 on an uncanny valley of voices. The test stimuli are part of the Perceptual Voice Qualities Database (PVQD; Walden, 2022). The set of voices consisted of 12 stimuli total: four being typical human voices, four being naturally pathological voices with the highest subjective severity ratings, and four being voices from the human voice database whose fundamental frequency was distorted by multiplication with 1000 using the STRAIGHT software (Kawahara et al., 2008). All voice stimuli consisted of individuals of either gender saying the sentences “the blue spot is on the key again. How hard did he hit him?”, and were 4 seconds in length.

Written text stimuli. Written text stimuli were taken from Chapter 5 on the uncanniness of distorted written text in familiar and unfamiliar languages. The sentences were either in English (familiar language and script), Icelandic (unfamiliar language, familiar script), or Babylonian Cuneiform (unfamiliar language and script). Distortions were created by moving and rotating the positions of the letters without changing their sequence in the word. Sentences were taken from the *Epic of Gilgamesh* provided by the Electronic Text Corpus of Sumerian Literature (ETCSL).

A summary of stimulus manipulations (except bodies and voices) is shown in *Figures 10.1* and *10.2*.

Figure 10.1

Face stimulus manipulations across conditions, divided by face type (rows) and distortion levels (columns).

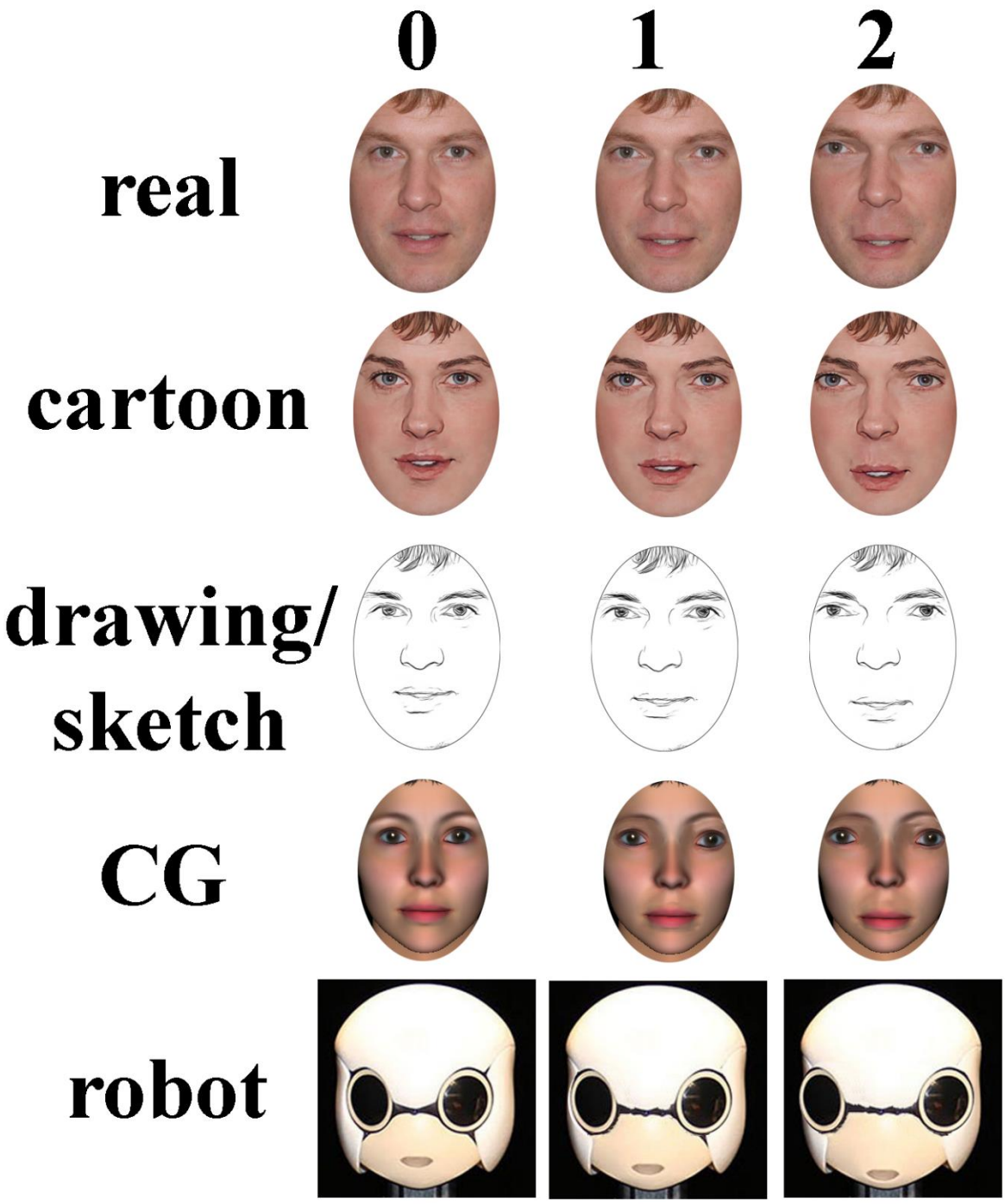
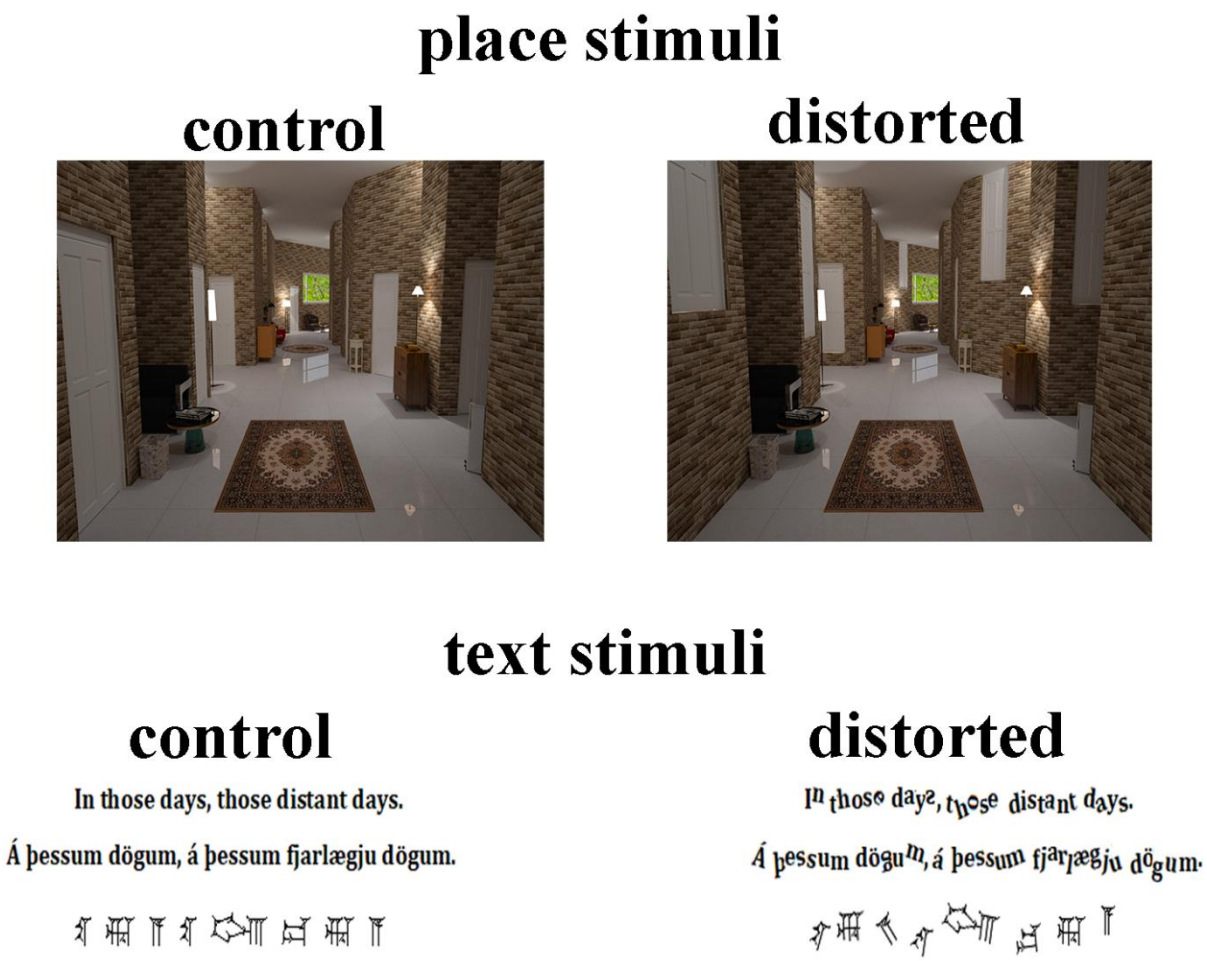


Figure 10.2

Place and word stimuli divided by stimulus condition. For place stimuli, a control and misplacement distortion is shown (see Chapter 6). For text stimuli, control and distorted

example text are shown for the three languages (top to bottom: English, Icelandic, Cuneiform).



Procedure

The study was conducted online. After participants received study information and gave informed consent, they completed the questionnaires. Afterwards participants were linked to the stimulus rating experiment which consisted of five sub-tasks: In the first, participants rated the face stimuli on the 0 – 100 scales on how eerie, strange, and humanlike they perceived the faces. For the following body, place, voice, and written text rating sub-tasks, participants rated stimuli on how 0 – 100 scales on how eerie and strange they perceived the stimuli (human likeness questions were omitted as some of the stimuli were not expected to vary on a human likeness dimension). Within each sub-task, stimuli were presented in a

randomized order, and participants had unlimited time to decide on each rating scale that was presented simultaneously with the stimulus. Participants received a short debrief after finishing the rating task.

Data analysis and availability

Data analysis was conducted via RStudio. Linear mixed models were used to investigate stimulus type effects and individual difference effects as they account for stimulus and participant effects. Outlier removal (1.5 IQR from median) was conducted for each rating scale and for each stimulus. For face stimuli, 109 uncanniness scores and 267 human likeness scores have been removed. For body, place, voice and text, stimuli, 2, 45, 25, and 22 uncanniness scores have been removed. Data, analysis, and non-copyrighted stimuli are available at https://osf.io/rz8d5.

Ethics statement

The study was conducted in alignment with the Declaration of Helsinki and approved by the Cardiff University Ethics Committee Board (EC.23.01.10.6716).

Results

Across sub-tasks, eerie and strange ratings were combined into an uncanniness rating by calculating the means on a trial-level. Scale correlations and Cronbach's alphas for each sub-task are shown in *Table 10.1*.

Table 10.1

Correlations of eerie and strange scales and Cronbach's alphas of the combined uncanniness ratings across sub-tasks.

Sub-task	Eerie-strange correlation	Cronbach's alpha
Face stimulus rating	0.74	0.85

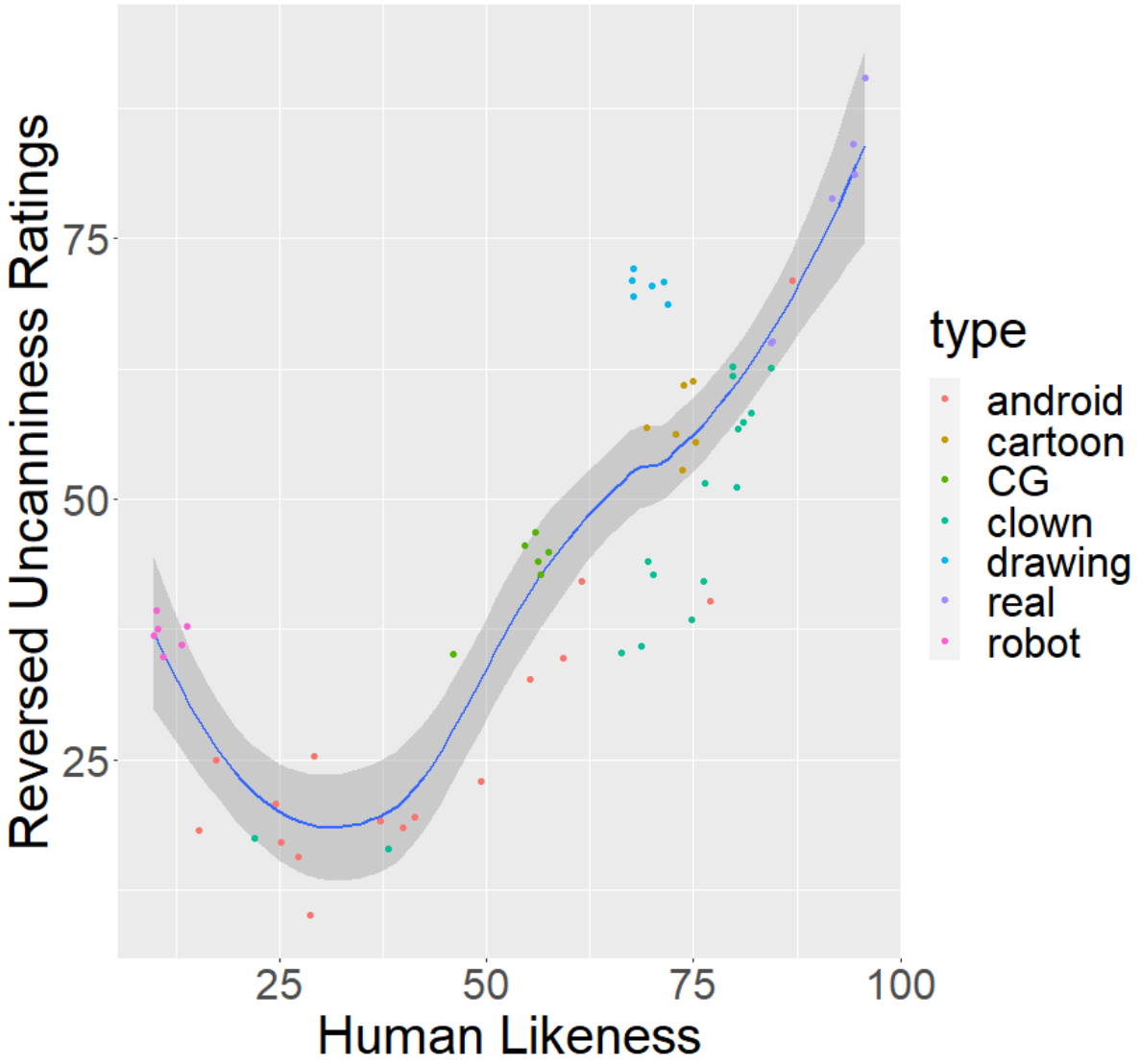
Body stimulus rating	0.74	0.85
Place stimulus rating	0.74	0.85
Voice stimulus rating	0.82	0.9
Text stimulus rating	0.63	0.77

Uncanny valley

Linear mixed models with linear, quadratic, and cubic functions of human likeness as fixed factors and stimulus and participant as random factors were conducted to investigate an uncanny valley. Results show that quadratic function could explain the data better than a linear function (χ^2 = 143.54, p < .001), yet the quadratic function was a better fit than a cubi function (χ^2 = 203.58, p < .001). Thus, a quadratic function of human likeness could best explain uncanniness (t(1136) = -12.04 p < .001, R^2_{cor} = 0.54, AIC = 62060). The data can be seen in *Figure 10.3*.

Figure 10.3

Human likeness ratings plotted against reversed uncanniness ratings across all face stimuli. The depicted plot is akin to an uncanny valley (Mori, 2012). Dots indicate stimuli, and grey areas show standard errors.



Uncanniness effects across stimulus types

Uncanniness effects were investigated using linear mixed models with distortion as fixed factors and participant and stimulus as random factors. If relevant to the hypotheses, additional variables like language familiarity and face type were added as additional fixed factors.

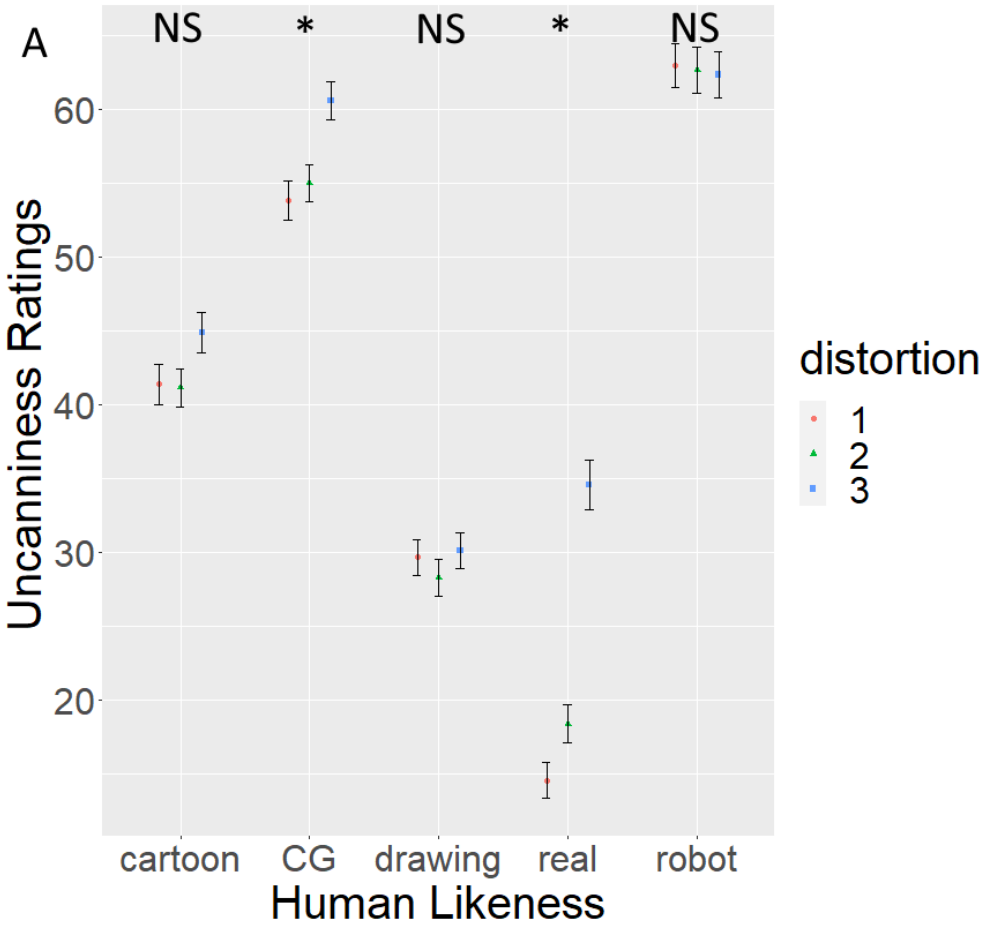
Face distortion. Within-subject and within-stimulus ANOVA with distortion and face type as factors showed significant main effects of face type ($F(4,123) = 95.12$, $p < .001$), distortion ($F(1, 120) = 21.34$, $p < .001$), and a significant interaction ($F(4, 120) = 3.4$, $p = .009$). To test whether the effect of distortion on uncanniness was stronger for more realistic faces, post-hoc

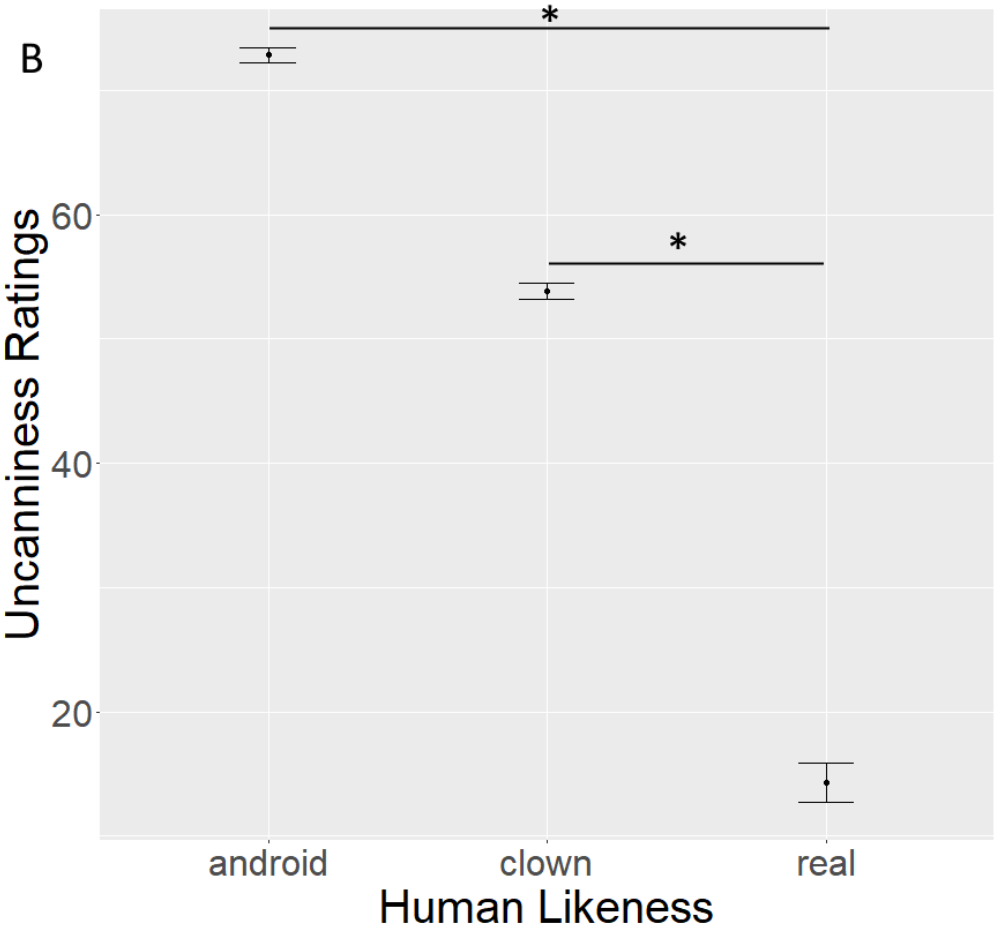
comparisons with Bonferroni-adjusted *p*-values were calculated between undistorted and max-distorted stimuli for each face type. Distortion significantly increased uncanniness in real ($t(3546) = -9.15$, $p_{adj} < .001$) and CG faces ($t(3546) = -3.08$, $p_{adj} = .005$), but did not affect cartoon ($t(3546) = -1.42$, $p_{adj} < .39$), drawing ($t(3546) = -0.35$, $p_{adj} = 1$), and robot faces ($t(3546) = 0.3$, $p_{adj} = 1$). Thus, the effect of distortion on uncanniness was present for real and CG faces, but not for (less realistic) drawing, cartoon, and robot faces. Results are shown in *Figure 10.4A*.

Android and clown stimuli. Within-subject ANOVA with stimulus type has been conducted using only undistorted human, android, and clown stimuli. Results reveal a significant main effect of stimulus type ($F(2,121) = 169.6$, $p < .001$). Post-hoc tests with Bonferroni-adjusted p-values furthermore revealed that android ($t(3987) = 031.89$, $p_{adj} < .001$) and clown stimuli ($t(3987) = 21.75$, $p_{adj} < .001$) were more uncanny than human stimuli. Results are shown in *Figure 10.4B*.

Figure 10.4

Differences between face type conditions and distortions. Figure 10.4A shows average uncanniness ratings across distortion levels for cartoon, CG, drawing, real, and robot faces. Figure 10.4B shows average uncanniness ratings for android, clown, and undistorted human faces. Error bars indicate standard errors.


A
NS
*
NS
*
NS
Uncanniness Ratings
60
50
40
30
20
distortion
1
2
3
cartoon
CG
drawing
real
robot
Human Likeness


B
*
*
Uncanniness Ratings
60
40
20
android
clown
real
Human Likeness

Body distortion. Within-participant, within-stimulus ANOVA with body distortion as fixed factor and stimulus and participants as random factors revealed a significant main effect of distortion ($F(2,121) = 40.04$, $p < .001$). Post-hoc comparisons with Bonferroni-adjusted p-values showed that distortions increased uncanniness from level 0 to 1 ($t(1442) = -8.01$, $p_{adj} < .001$), 0 to 2 ($t(1442) = -16.98$, $p_{adj} < .001$), and 1 to 2 ($t(1442) = -8.96$, $p_{adj} < .001$). Data is shown in *Figure 10.5A*.

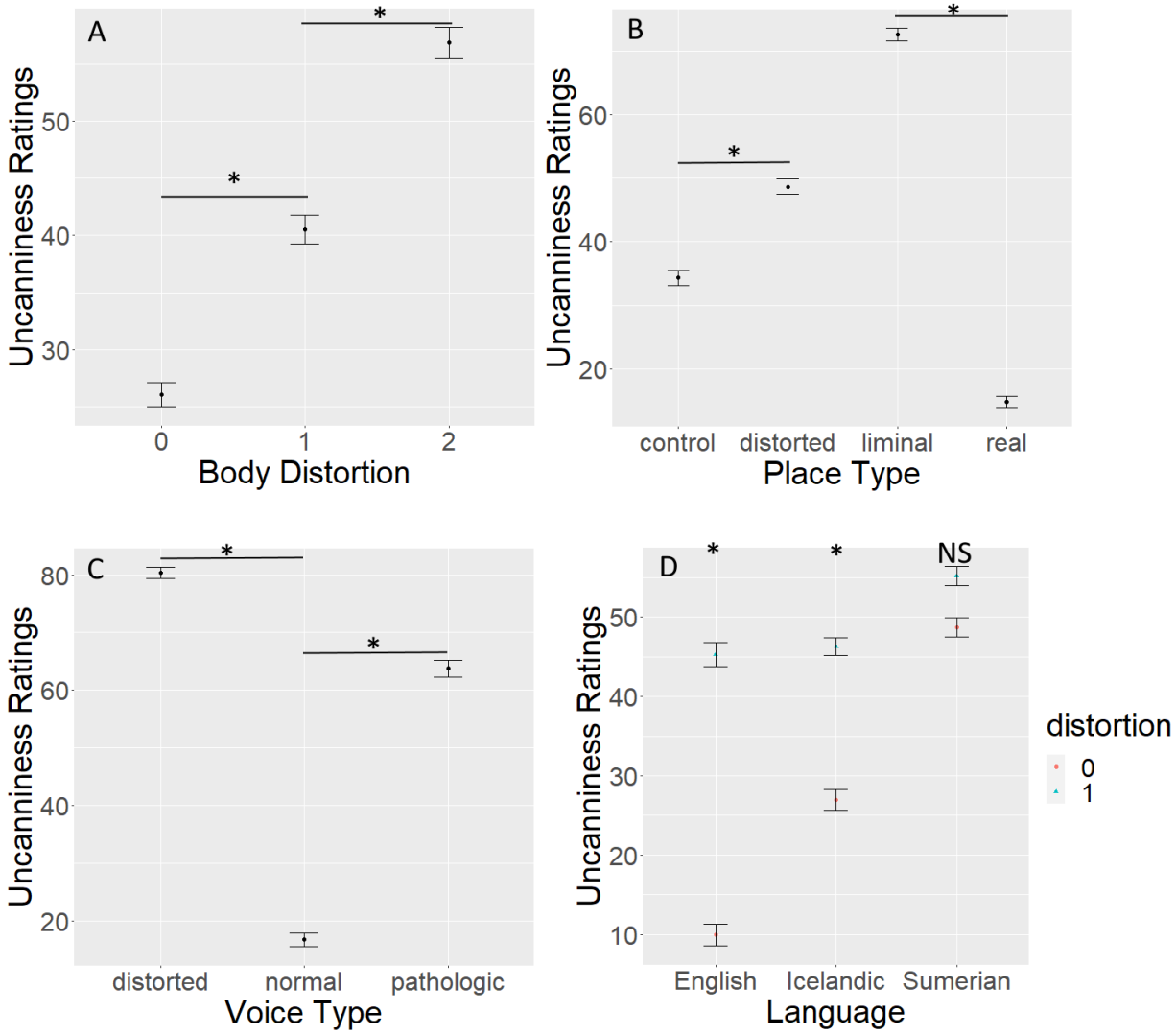
Place distortion. For uncanniness in places, within-subject within-stimulus ANOVA with place type as fixed factor and stimulus and participants as random factors revealed a significant main effect of place type ($F(3,120) = 152.8$, $p < .001$. Bonferroni-adjusted post-hoc tests showed that distorted places were more uncanny than their non-distorted paired counterparts ($t(2115) = -9.349$, $p_{adj} < .001$), and natural distorted places were more uncanny than natural real places ($t(2115) = 41.278$, $p_{adj} < .001$). Data is shown in *Figure 10.5B.*

Voice distortion. Within-subject within-stimulus ANOVA with voice type as fixed factor and stimulus and participants as random factors revealed a significant main effect of voice type ($F(2,121) = 152.8$, $p < .001$). Bonferroni-adjusted post-hoc tests revealed that distorted voices ($t(1054) = 40.76$, $p_{adj} < .001$) and pathological voices ($t(1054) = 30.341$, $p_{adj} < .001$) were more uncanny than typical voices. Data is shown in *Figure 10.5C.*

Text distortion. Within-subject within-stimulus ANOVA with text distortion and language as fixed factors and stimulus and participants as random factors revealed significant main effects of language ($F(2,121) = 42.25$, $p < .001$), distortion ($F(1,122) = 114.11$, $p < .001$), and an interaction between these factors ($F(2,121) = 9.8$, $p < .001$). Bonferroni-adjusted post-hoc tests revealed that for English ($t(1421) = 15.21$, $p_{adj} < .001$) and Icelandic sentences ($t(1421) = 8.24$, $p_{adj} < .001$), distortion increased uncanniness, while it did not affect Babylonian sentences ($t(1421) = 2.73$, $p_{adj} = .069$). Data is shown in *Figure 10.5D.*

Figure 10.5

Average uncanniness ratings across stimulus types and conditions. Figure 10.5A shows uncanniness ratings across body distortion levels. 10.5B shows uncanniness ratings across place types. 10.5C shows uncanniness ratings across voice types. 10.5D shows uncanniness ratings across text language and distortion levels.



Summary of results

Validation analysis showed that, consistently across stimulus domains, incremental distortions increased uncanniness: This was the case for faces, bodies, places, voices, and written text. Furthermore, naturally occurring deviating stimuli were more uncanny than non-

deviating counterparts, as was the case for android and clown compared to human faces, deviating ("liminal") compared to typical places, and natural pathological voices compared to non-pathological voices. Finally, distortion effects were moderated by realism level in faces and language/script familiarity in written text stimuli, indicating that specialization with a stimulus category increases distortion sensitivity. In total, the results confirm that distortions can cause uncanniness across stimulus categories, and that this effect is more pronounced with higher familiarity or specialization. Thus, the validation hypotheses are confirmed.

Individual difference analysis

Individual differences were for each construct were calculated via means of item responses. Each questionnaire's Cronbach's alpha and intercorrelations between individual differences (including significance marks) are shown in *Table 10.2*.

Table 10.2

Cronbach's alphas and intercorrelations of individual difference measures.

Measure (Cronbach's alpha)	FCQ	DS-R	PDM	PNS	AN
FCQ (0.98))	1				
DS-R (0.79)	0.42	1			
PDA	0.09	0.27	1		
PNS (0.85)	0.12	0.33	0.31	1	
AN (0.91)	0.31	0.37	0.2	0.53	1

Note: FCQ = Fear of Clowns Questionnaire; DS-R = Disgust Scale-Revised; PDA = Pattern deviancy aversion; PNS = Personal Need for Structure; AN = Anxiety-Neuroticism

Within-participant within-stimuli ANOVAs have been conducted for each relevant analysis. For face stimuli, stimuli were divided by stimulus type including all distortion levels and effects of individual difference measures on uncanniness ratings were analysed for each type separately. For body stimuli, interactions between distortion level and individual differences were investigated. For place data, interaction between place type and individual difference measures were investigated. For voice data, distorted and pathological voices were separated and interactions between voice type and individual difference measures were investigated. For word data, interaction effects between distortion and individual difference measures were investigated across languages. The results are summarized in *Table 10.3*.

Table 10.3

Test statistics across stimulus type and individual difference measure. For each stimulus type, within-subject ANOVAs were conducted with individual differences as combined factors. "n.s." indicates that no significant effect was obtained.

Stimulus	Fear of clowns	Anxiety-Neuroticism	Deviancy aversion	Need for structure	Disgust sensitivity
Human	$F(1,122) = 10.81$, $p = .001$	n.s.	$F(1,122) = 10.57$, $p = .001$	n.s.	n.s.
Cartoon	$F(1,122) = 8.19$, $p = .004$	n.s.	n.s.	n.s.	n.s.
Drawing	$F(1,122) = 42.41$, $p < .001$	n.s.	n.s.	$F(1,122) = 3.97$, $p = .047$	n.s.
CG	$F(1,122) = 4.03$, $p = .045$	$F(1,122) = 11.17$, $p > .001$	n.s.	n.s.	$F(1,122) = 9.63$, $p = .002$

Robot	$F(1,122)$ = 12.69, $p > .001$	$F(1,122)$ = 22.58, $p > .001$	$F(1,122) = 5.8$, $p = .016$	$F(1,122)$ = 6.49, $p = .011$	n.s.
Android	n.s.	$F(1,122)$ = 9.52, $p = .002$	$F(1,122)$ = 10.03, $p = .002$	n.s.	$F(1,122)$ = 5.6, $p = .018$
Clown	$F(1,122)$ = 79.22, $p > .001$	n.s.	$F(1,122)$ = 5.84, $p = .016$	n.s.	$F(1,122)$ = 5.92, $p = .015$
Bodies	$F(1,122)$ = 4.29, $p = .038$	n.s.	n.s.	n.s.	n.s.
Places	n.s.	n.s.	$F(1,122)$ = 4.13, $p = .006$	n.s.	n.s.
Distorted voices	$F(1,122)$ = 15.09, $p > .001$	n.s.	$F(1,122)$ = 10.03, $p = .001$	n.s.	n.s.
Path. voices	n.s.	n.s.	$F(1,122)$ = 7.86, $p = .005$	n.s.	n.s.
Written text	n.s.	n.s.	n.s.	n.s.	n.s.

Discussion

Throughout multiple tasks involving stimuli of different domains, the role of individual differences on uncanniness effects caused by deviations were investigated. First, the validation of the uncanny valley and uncanniness effects across stimulus conditions are discussed, followed by a discussion of the individual difference results and finally a discussion on the results' implications for the understanding of the uncanny valley.

Validation of uncanniness effects

Uncanny valley. A quadratic relationship of human likeness of different face types could best explain uncanniness ratings, akin to an uncanny valley (Mori, 2012). Incremental facial distortions increased uncanniness more in more realistic faces (e.g., real human versus robot faces), and android and clown stimuli were more uncanny than normal human stimuli. Results confirm the uncanny valley effect (Mori, 2012) and the moderating role of face realism on distortions (Chapters 2 and 3). Furthermore, clowns were rated as more uncanny than humans, indicating that they may fall into an uncanny valley (Tyson et al., 2023). Thus, the validation hypothesis on the uncanny valley is confirmed (*uncanny valley hypothesis*), just as the hypothesis predicting an increase of uncanniness across face distortions moderated by face realism (*face uncanniness hypothesis*).

Uncanniness effects. Across stimulus conditions, distortions increased uncanniness ratings: Incremental distortion of body part lengths increased uncanniness; manipulated or naturally distorted physical places were more uncanny than controls; artificially distorted or naturally pathological voices were more uncanny than healthy undistorted voices; Finally, language familiarity moderated the effect of orthographic configural distortion of written text on uncanniness, with the effect of distortions being stronger in familiar versus unfamiliar languages. Thus, uncanniness effects were confirmed across stimulus categories, and all further validation hypotheses are confirmed (namely *body uncanniness hypothesis*, *place uncanniness hypothesis*, *voice uncanniness hypothesis*, *written text uncanniness hypothesis*). In addition, as clown stimuli were more uncanny than human stimuli, the first coulrophobia hypothesis was confirmed.

Effect of individual difference variables

Five individual difference variables (Anxiety-Neuroticism, coulrophobia, deviancy aversion, disgust sensitivity, and need for structure) were used as predictors of uncanniness or the uncanny valley effect.

Across face types, uncanniness of distorted faces was explained by individual differences in fear of clowns. The role of other individual differences in the uncanniness of distorted faces differed across types of faces: For human and robot faces, deviancy aversion best explained uncanniness. For CG and robot faces, anxiety-neuroticism explain uncanniness. For drawing and CG faces, need for structure contributed to uncanniness, and for CG faces, disgust sensitivity explained uncanniness.

For android faces, anxiety, deviancy aversion, and disgust sensitivity predicted uncanniness ratings. For clown faces, coulrophobia, deviancy aversion, and disgust sensitivity predicted uncanniness ratings. Thus, the roles of individual differences on uncanniness mostly differed depending on the stimulus type. However, for androids (which are the ecologically most relevant stimuli here), contributing roles of anxiety, deviancy aversion, and disgust sensitivity were found.

For body stimuli, only coulrophobia predicted uncanniness. For place stimuli, only deviancy aversion predicted uncanniness. For voice stimuli, deviancy aversion predicted the uncanniness of pathological and distorted voices, and coulrophobia for distorted voices. Finally, for written text stimuli, no individual difference variable was a significant predictor. Thus, uncanniness effect hypotheses are mostly confirmed for deviancy aversion and partially confirmed for coulrophobia. No evidence was found for the other three variables.

Discussion of individual difference variables

Anxiety-Neuroticism. The anxiety sub-facet of neuroticism was unable to predict uncanniness ratings in any of the analyses. Present results confirm the findings by MacDorman and Entezari (2015) who found that anxiety-neuroticism significantly predicted uncanniness ratings of androids: In this study, anxiety-neuroticism also predicted android uncanniness. Furthermore, anxiety levels also predicted the uncanniness of distorted CG and robot faces, but did not affect the uncanniness of more realistic faces, clowns, or uncanniness effects of distorted non-face stimuli.

Coulrophobia. Individual differences in self-reported fear of clowns significantly predicted the uncanniness of clowns and facial distortions across all levels of face realism. Furthermore, Coulrophobia predicted uncanniness of distorted bodies (incrementally elongated or shortened body limbs) and distorted voices. Meanwhile, coulrophobia was not associated with the uncanniness of androids, pathological voices, distorted places, or orthographically distorted written text.

Increased uncanniness of clowns in individuals with higher coulrophobia fits the prediction that coulrophobia is associated with the uncanny valley: Tyson et al. (2023) found that participants who reported higher levels of coulrophobia also reported that clowns look disturbing or out of place, which as interpreted as happening due to an uncanny valley effect. Further evidence for the connection between coulrophobia and the uncanny valley is the observation that uncanniness created by incrementally distorted faces was also explained by individual differences in coulrophobia. Coulrophobia predicting both clown and distorted face uncanniness suggests that whatever sensitizes individuals to coulrophobia also sensitizes them to the uncanniness of face distortions. As coulrophobia also predicted body and voice uncanniness, the dislike of distortion or exaggeration of humanlike features may be such a mechanism. Given that clowns tend to wear costumes exaggerating body proportions (e.g.,

big red noses or long clown shoes) the dislike of such features may be what is driving the effect of coulrophobia on the uncanniness of distorted faces, clowns, and distorted bodies and voices.

The current study provides interesting insights on the contributors to coulrophobia: Firstly, its association with android uncanniness suggest a connection with the uncanny valley effect. Second, given that an association between coulrophobia and uncanniness was observed across multiple categories related to human appearance and sound, the aetiological mechanisms underlying coulrophobia go beyond clown stimuli but extend to all biologically human stimuli, and may encompass a sensitivity and dislike towards deviating biological features in general. Alternatively, specific aetiological mechanisms that lead to coulrophobia (e.g., media exposure; Tyson et al., 2023) may spill-over to categories beyond clowns in the form of a higher sensitivity to distorted biological stimuli.

It should be noted that this study used a fear of clown questionnaire developed to measure individual expressions of coulrophobia within the normal population, which is not suitable to diagnose clinical levels of coulrophobia. Thus, generalizations of the results or interpretations onto individuals with clinically significant expressions of coulrophobia are not warranted.

Deviancy aversion. Deviancy aversion describes an individual's tendency to dislike deviations in simple patterns, and can be generalized onto the dislike of more complex deviancy such as statistical minorities (Gollwitzer et al., 2017). In this study, deviancy aversion predicted the uncanniness of distorted human, android, and clown faces, in addition to distorted bodies and distorted and pathological voices. As stimuli with the uncanny valley are marked with near humanlike, "not quite right" characteristics (Diel et al., 2022; Mori, 2012), uncanniness may stem from deviancy aversion in deviations from typical human appearance and behaviour. Just as simple pattern deviancy aversion can predict the dislike of

social pattern deviancy (e.g., statistical minorities; Gollwitzer et al., 2017; Gollwitzer et al., 2020) by measuring a domain-general dislike of pattern violation, so could this domain-general mechanisms also explain the dislike of stimuli that violate human appearance, like androids, clowns, or distorted faces. Deviancy aversion could explain uncanniness of clowns even when coulrophobia was controlled for, indicating that deviancy aversion affects clown uncanniness even in individuals who do not report a fear of clowns.

Beyond humanlike stimuli, deviancy aversion also predicted the uncanniness of distorted places and distorted and pathological voices, supporting the notion of a domain-general mechanism: Physical places marked by unusual feature combinations, such as a lack of, misplaced, or repeated features, would deviate from the typical appearance of physical places (Chapter 6). Individuals with higher deviancy aversion expectedly show a greater dislike of such places.

The results in voice sensitivity show the first evidence of a cross-modal (visual-auditory) predictive effect of deviancy aversion: While the deviancy aversion measure relies on visual stimuli, it predicted uncanniness ratings of auditory stimuli. These results indicate that deviancy aversion and its underlying mechanisms may show the same or analogous processes across modalities.

The association between deviancy aversion and uncanniness of pathological voices corresponds to previous research finding that deviancy aversion predicts dislike of individuals with disabilities (Gollwitzer et al., 2020). As pathological voices are marked by atypical expressions in vocal dimensions like formant frequencies (Davis, 1979), deviancy aversion differences may sensitize to such atypicalities in voices. As disfigured faces have been shown to elicit responses akin to an uncanny valley (Diel & MacDorman, 2021),

deviancy aversion may explain this devaluation of biologically non-typical appearance or behaviour.

Individuals with higher expressions of deviancy aversion and an uncanny valley effect may be more sensitive to effects of processing disfluency (Winkielman et al., 2003): deviating patterns, be they simple or more complex such as faces, may need additional processing power, which decreases aesthetic evaluation. Alternatively, Deviancy aversion may be linked to differences in the processing of expectation violations in predictive coding (Friston, 2010), which has been associated with the uncanny valley in past research (e.g., Saygin et al., 2012). Future research may look into whether individual expressions in deviancy aversion predict the strength of neurophysiological responses to deviating stimuli which may be associated with processing disfluency or prediction errors.

Disgust sensitivity. Disgust responses as warnings against contamination may contribute to the dislike of individuals deviating from typical biological appearance (Park, Faulkner, & Schaller, 2003), and androids may be uncanny because they activate such processes (MacDorman & Ishiguro, 2006; Moosa & Ud-Dean, 2010). The emotion of disgust has been associated with the uncanny valley in the past (Ho et al., 2008), and disgust sensitivity has been found to sensitize uncanniness in androids (MacDorman & Entezari, 2015). In this study, disgust sensitivity has been found to predict the uncanniness of distorted CG faces in addition to clown and android faces. Thus, disease avoidance explanations of the uncanny valley have found support in the current results.

Previous research on coulrophobia found that people with self-reported fear of clowns also tended to report disgust-related response towards clowns (Tyson et al., 2023; see also Planting, Koopowitz, & Stein, 2022). In accordance, coulrophobia here was correlated with disgust sensitivity (r = 0.42) and disgust sensitivity, in addition to coulrophobia and deviancy

aversion, predicted clown uncanniness. It is possible that clowns elicit disgust responses for similar reason as to why disgust is thought o be associated with the uncanny valley: entities appearing human yet deviating from human appearance may trigger disease detection mechanisms including disgust responses which would motivate the individual to dislike and avoid the trigger (MacDorman & Ishiguro, 2006). Alternatively, as clowns are known to show behaviour that break social taboos and norms, disgust responses may be related to the violation of social norms (Chapman & Anderson, 2012). In any case, the current results support the notion that disgust responses are related to coulrophobia, and the dislike of clowns in general.

Need for structure. Individual differences in the personal need for cognitive structures have been associated with the uncanny valley in past research (Lischetzke et al., 2017). However, these results were not replicated in this study. Instead, need for structure only predicted uncanniness of distorted drawing and robot faces, which – given their low realism or human likeness levels – are not typically associated with the uncanny valley. Given the correlations with other variables, other intercorrelating individual differences may have explained uncanniness better than need for structure, notably disgust sensitivity ($r = 0.33$) and deviancy aversion ($r = 0.31$).

Alternatively, need for structure may be more associated with violations of schematic representations rather than highly realistic instances. Both drawing and robot faces show human face patterns on a comparatively abstract level. Personal need for structure may be related to violations of such abstract cognitive schemata rather than violations of more realistic, specific instances of faces. This would explain why need for structure was only relevant for uncanniness in low-realism face types, but irrelevant in highly realistic faces or androids.

Heterogeneity of the uncanny valley

The results indicate that stimuli may be uncanny for different reasons. While some individual difference variables showed surprisingly consistent effects on uncanniness across stimulus categories and modalities (e.g., deviancy aversion), no variable could consistently explain uncanniness across all stimulus types – although out of coulrophobia, deviancy aversion, and disgust sensitivity showed relevance throughout multiple stimulus categories, including androids.

Different predictors for different stimulus categories may underlie different processing mechanisms. For example, while coulrophobia may be related to being disturbed by exaggerated humanlike stimuli and disgust sensitivity may be related to indicators of disease or norm violations. Meanwhile, deviancy aversion may explain domain-independent dislike of pattern violations while need for structure may relate to violations of cognitive schemata like abstract representations of human faces, or conceptual violations of categories caused by morphing (Lischetzke et al., 2017). Finally, anxiety may facilitate stressful responses caused by stimuli already perceived as uncanny (MacDorman & Entezari, 2015). Stimuli relevant to the uncanny valley, like androids, may elicit multiple mechanisms (e.g., deviancy aversion and disgust sensitivity) that cumulate to strong negative responses. Finally, different processing mechanisms may elicit slightly different negative subjective experiences. However, direct or indirect measures may lack sufficient discriminability for negative subjective experience, which are instead summarized into, for example higher uncanniness ratings.

Chapters 2 to 10 established the foundation of a refined theory of the uncanny valley. However, barely any of the research used android stimuli (except for Chapters 8 and 10). Chapter 11 presents research supporting the refined theory to explain uncnaniness of an humanlike android.

Chapter 11: Configural processing enhances the uncanniness of distorted dynamic facial expressions

Methods, experiments, and large portions of the introduction and discussion in this chapter is currently published in the journal *Frontiers in Psychology* and in review in the journal *BMC Research Notes*.

Introduction

Previous chapters developed and tested the refined theory of the uncanny. Most stimuli used in previous chapters however relied on image manipulation, which may not represent ecologically relevant uncanny stimuli. In Chapter 11, the refined theory is applied to the dynamic facial expressions of a realistic android.

Several empirical studies have investigated the relationship between human likeliness and emotional responses to humanlike artificial entities. Although the results are not completely consistent across studies (Diel et al., 2022), a recent meta-analysis work analyzed the data from 49 studies that investigated the relationship between human likeness and likeability to robot agents (Mara et al., 2022). Researchers identified a cubic, sigma-shaped function that reflects the relationship between the human likeness and likability of multiple artificial entities. These results suggest that the relationship between human likeness and the emotional impressions of artificial entities takes on a non-linear shape, which might be associated with the uncanny valley phenomenon.

However, uncertainties remain concerning this sigma-shaped relationship between likability and anthropomorphism. First, a meta-analysis by Mara et al. (2022) focused on Godspeed likability scales (Bartneck et al., 2009) and non-realistic humanlike robots (e.g., NAO), whereas uncanny valley research typically assesses uncanniness and more realistic robots or computer-generated (CG) characters as well as complete human stimuli (Diel et al., 2022).

Furthermore, the psychological mechanisms underpinning the cubic relationship between human likeness and emotional impressions remain unknown. We hypothesize that the candidate mechanisms might include the configural processing of faces and facial expressions. Perhaps the uncanny valley effect is rooted in enhanced error processing in specialized categories (Chapters 2 to 4; Chattopadhyay & MacDorman, 2017; Diel & MacDorman, 2021; Kätsyri, de Gelder, & Takala, 2019; MacDorman et al., 2009). Atypicalities or deviations might induce negative aesthetic evaluations, which may be especially sensitive in stimulus categories that elicit specialized processing. Such specialized (in this case configural) processing of faces depends on an upright facial orientation, and a global inversion of faces disrupts this processing style (*inversion effect*; Carbon & Leder, 2006; Kanwisher & Moscovitch, 2000). Configural processing, which also improves the processing of facial expressions, is disrupted when expressions are inverted (Ambadar, Schooler, & Cohn, 2005; Bould & Morris, 2008; Tobin, Favelle, & Palermo, 2016). The variance of facial aesthetic ratings falls when faces are presented in an inverted manner, most likely because the accuracy of face processing falls (Bäuml, 1994; Leder, Goller, Forster, Schlageter, & Paul, 2017; Santos & Young, 2008). Furthermore, the uncanniness ratings of faces are less severe when faces are presented in an inverted manner due to a decreased ability to detect changes or distortions in them (Chapter 2). However, the effect of inversion on likability ratings has not yet been investigated across a range of entities that have different human likenesses or anthropomorphic qualities; neither have the inversion effects on uncanniness been tested for dynamic facial expressions. Since specialized processing is more pronounced in more realistic faces (Crookes et al., 2015), inversion may disrupt more subtle differences in the aesthetic ratings of highly realistic dynamic expressions. This effect of specialized processing on dynamic face processing might explain why subtle facial movements in realistic androids sometimes appear eerie or uncanny.

Experiment 16

The purpose of this work is to investigate whether a cubic relationship between human likeness and uncanny ratings can be shown for humanlike agents and whether configural processing plays a role in such a relationship. To test the humanlike agents, we presented dynamic emotional expressions of three kinds of faces: human, android, and CG. We tested the effect of configural processing of faces by comparing the upright and inverted presentations of the facial stimuli. We presented various types of facial stimuli using the emotional facial expressions of negative and positive valence (i.e., anger and happiness) and facial expressions with different facial action patterns over time, which elicited slightly different emotional impressions (Diel et al., in review). In accordance with previous meta-analyses (Diel et al., 2022; Mara et al., 2022), we investigated the uncanny valley effect by testing for a cubic function between the ratings of aesthetics and human likeness:

1. Hypothesis 1: A cubic function explains the relationship between uncanniness and human likeness in upright facial expressions.

However, since the non-linear relationships between uncanniness and human likeness may occur due to specialized processing of faces and facial expressions, this effect should not occur when the presented expressions are inverted:

2. Hypothesis 2: A linear function explains the relationship between uncanniness and human likeness in inverted facial expressions.

No differences in the nature of the functions of uncanniness for upright and inverted expressions would indicate a lack of inversion effect, suggesting that specialized processing plays little or no role in the evaluation of the aesthetics of artificial entities.

Furthermore, the effect of asynchronies on the uncanniness of facial expressions is

investigated, and whether this effect is influenced by specialized processing. . First, deviations or atypicalities are thought to increase uncanniness especially in categories that elicit specialized processing. Asynchrony is here understood as a form of temporal deviation from a dynamic pattern of face muscle motion. Second, an orientation effect is investigated as the degree of specialization is thought to moderate the effect of uncanniness. Finally, an actor effect is investigated as the degree of specialized processing should be increased for more realistic (embodied) actors compared to computer-generated ones.

1. Asynchronous motion in facial expression increases uncanniness (*asynchrony effect*)
2. Inversion reduces the effect of asynchrony on uncanniness (*uncanniness inversion effect*)
3. The uncanniness inversion effect is present for human and android but not CG expressions (*actor effect*)

Methods

Participants

Sixty-four Japanese volunteers participated in this study (31 females, 31 males, and two who preferred not to specify; mean ± SD age, 30.65 ± 3.88 years). The required sample size was determined using an a priori power analysis with G*Power software ver. 3.1.9.2 (Faul, Erdfelder, Lang, & Buchner, 2007). As an approximation of the present analysis using linear mixed-effects models containing seven dependent variables (i.e., the interaction model), a multiple linear regression model with seven dependent variables was analyzed. A power analysis for the coefficient evaluation (two-tailed) with the assumption of Cohen's *f* of 0.15 (medium size effect), α level of 0.05, and power (1 - β) of 0.80 concluded that 55 participants were needed. Participants were recruited through web advertisements distributed by CrowdWorks (Tokyo, Japan). After the procedures were explained, all participants provided

written informed consent for joining the study, which was approved by the Ethics Committee of RIKEN. The experiment was performed in accordance with the Declaration of Helsinki.

Materials

Android. We used an android named Nikola, which was developed using 35 actuators to naturally simulate the relevant facial actions for recreating six basic human emotions (Sato et al., 2022). Since Nikola's pneumatic actuators have a temporal resolution of milliseconds, they can adequately manipulate the synchronous motions of natural emotion expressions.

Videos. Human videos were created using angry and happy expressions from the AIST Facial Expression Database (Fujimura & Umemura, 2018). Android videos were created by filming Nikola's frontal emotion expressions. CG videos were created using FACSGEN software (Krumhuber, Tamarit, Roesch, & Scherer, 2012; Roesch et al., 2011). For both the android and CG faces, the following face AUs were used for the expressions: angry: 4, 5, 7, 23; happy: 6, 12.

Asynchronies were either none (original video or synchronous motion), a 250-ms delay (face's upper right half moved with a 250-ms delay and upper left half with a 500-ms delay) and 500 ms delay (face's upper right half moved with a 500-ms delay and upper left half with a 1000-ms delay). The lower half of the face started to move at the same time in each condition. For the android videos, asynchronous motion was created by delaying the programmed motion onset. For the CG videos, asynchronous motion was first created by first delaying the programmed motion onset of the upper half and then the left upper half using the Adobe Premiere video editing tool. For the human videos, both the upper left-side and right-side asynchronies were created by delaying their onset using the same editing tool.

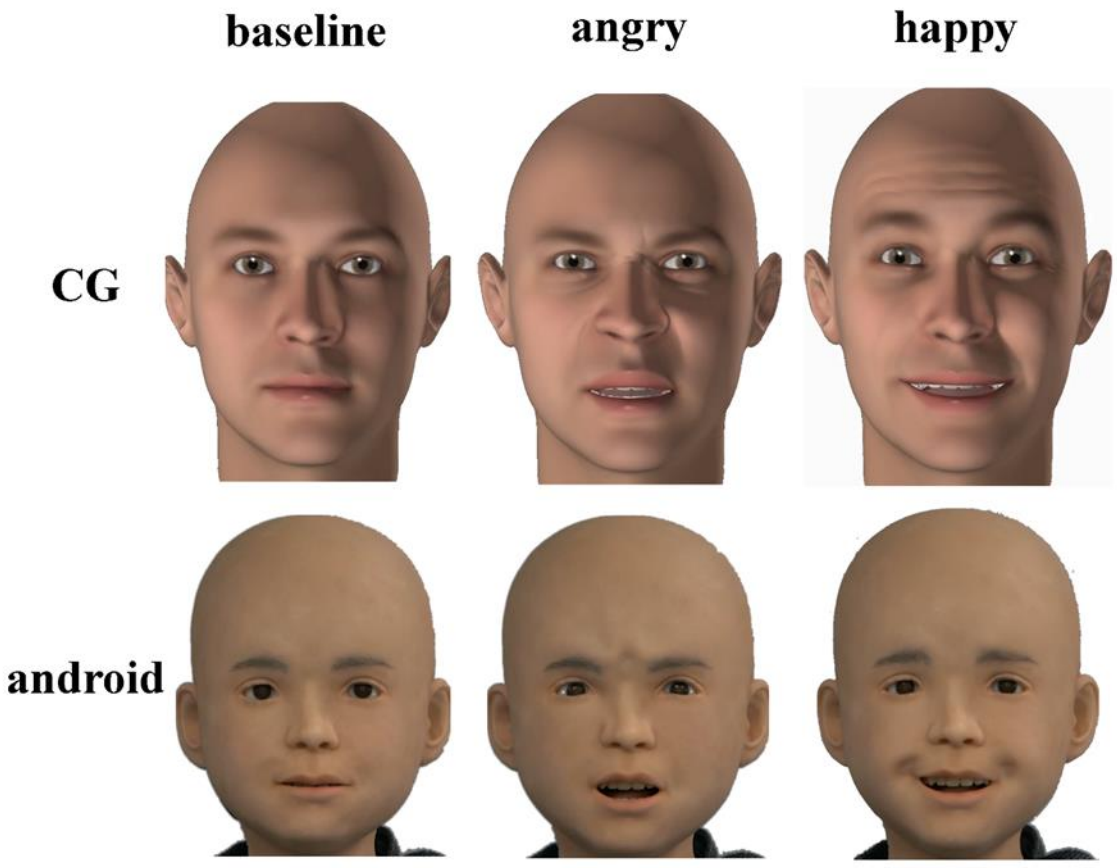
All the videos were edited so that the noses of each actor were at the same height, and they were cut at the neck (bottom), head (top), and ears (left and right). A white background was

displayed behind the actors. All videos were 1.25 seconds long and depicted the onset of one of two emotion expressions: angry or happy.

We used a total of 36 videos: 3 actors, 3 asynchrony levels, 2 orientations, and 2 emotions. Screenshots of the android and CG expressions are shown in *Figure 11.1*.

Figure 11.1

Stimuli across conditions. CG (top) and android (bottom) stimuli across emotion conditions: Baseline (neutral) expressions are to right, followed angry and happy expressions. Human expressions are not shown due to copyright issues.



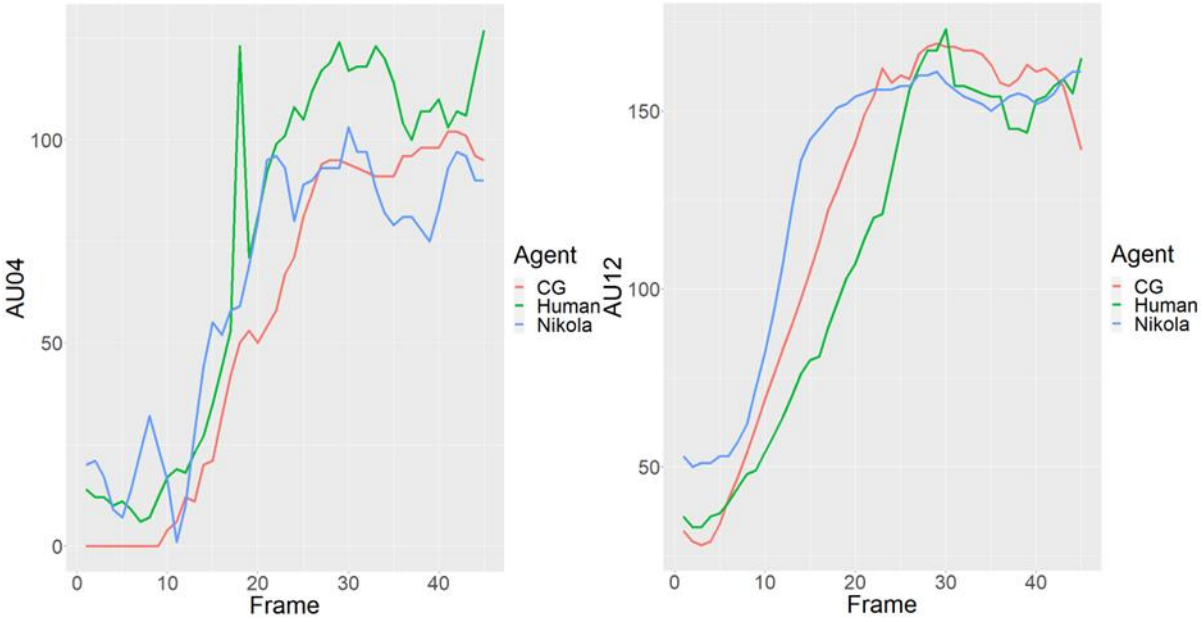
Stimulus validation.

A stimulus validation pilot study was conducted to test whether the objective and subjective emotion expressions differed among the three actors.

Objective expressions. For their validation, facial movements of the base stimuli of angry and happy expressions for each actor were analyzed using OpenFace. AU4 and AU12 values indicated angry and happy expressions. The trajectories of both the AUs are depicted in *Figure 11.2* and indicate no strong deviations among the three actors, showing that the intensity of the AU expressions are analogous across the actors.

Figure 11.2

AU intensities across conditions. Intensity of face action units AU4 and AU12 across actor type: Values were analyzed automatically using OpenFace.



Subjective expressions. For the subjective expressions, a questionnaire study was conducted. Based on the bipolar valence-arousal modes, single-scale items of valence and arousal were used for the assessment of the emotional expressions. Eleven participants rated the faces on the following scales ranging from 0 to 100: how angry is the face, how happy is it, its emotional arousal, and its emotional valence. The study was conducted online. Its results show no significant main effects of actor type on the ratings for how happy the faces are ($F(2,63) = 0.07$, $p = .93$) or how angry they are ($F(2,63) = 0.28$, $p = .76$); they also show no

significant main effects on arousal ($F(2,63) = 0.05$, $p = .95$) or valence ($F(2,63) = 0.12$, $p = .89$) ratings.

Thus, for both emotions, the indicators for both the objective and subjective emotional expressions did not differ across actors.

Procedure

The experiment was conducted online. After providing informed consent, participants were linked to the experiment page and shown each video in a randomized order. They rated each video on the three scales used in a previous study (Diel et al., 2022): *uncanny*, *strange*, and *humanlike*. They were shown the terms with a scale and selected how much the depicted video was uncanny/strange/humanlike on scales ranging from 0 to 100. Participants had unlimited time for each scale and were allowed to freely rewatch the videos at any time during the rating.

Statistical analysis

All the statistical analyses were performed using the statistics and machine learning toolbox in MATLAB 2020a (MathWorks, Natick, MA, USA). The relationships between the uncanny and humanlike ratings were analyzed based on the purpose of this study. The data, stimuli (except the human videos), and the analysis are available: https://osf.io/9cmhp.

Results

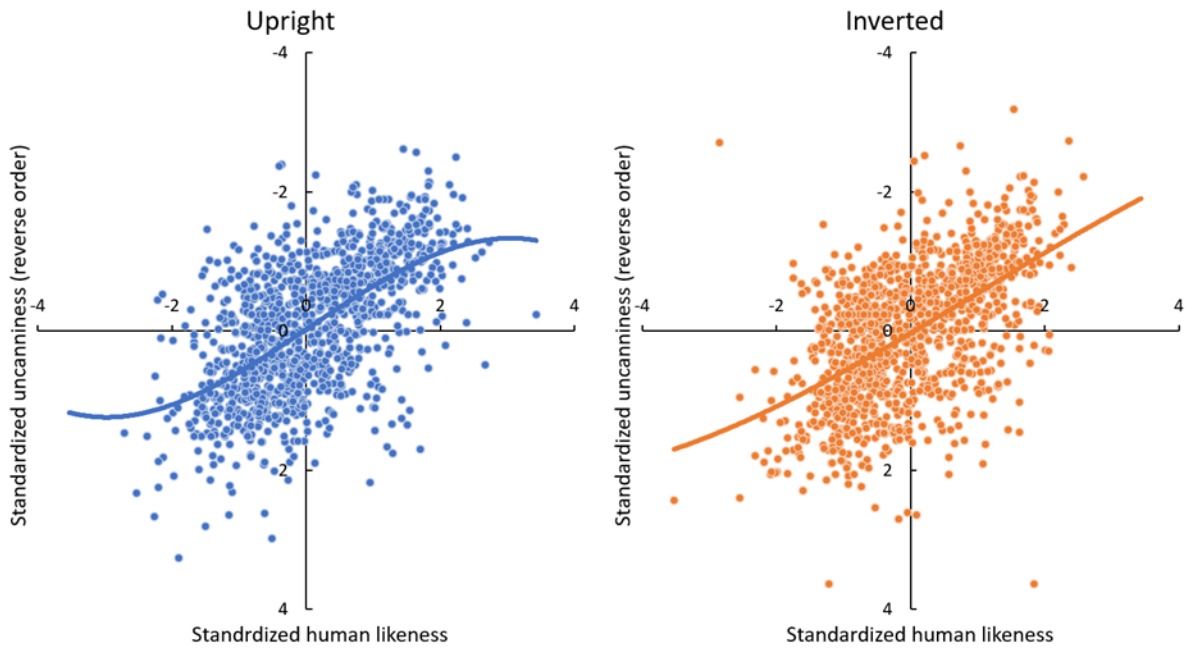
Inversion effect on polynomial function of human likeness and uncanniness

The relationships between the uncanny and humanlike ratings are illustrated in *Figure 11.3*.

Figure 11.3

Regression lines across orientation conditions. Scatter plots and cubic regression lines between uncanny and humanlike ratings for upright and inverted presented faces:

Standardized scores are shown to indicate consistent patterns across participants. Points represent raw data points.



Linear mixed-effect models were constructed using uncanny ratings as dependent variables. The main effect model included linear, quadratic, and cubic functions of human likeness and orientation as independent variables, and the interaction model included linear, quadratic, and cubic functions of human likeness, orientation, and interactions between each function of human likeness and orientation as independent variables. Random by-participants intercepts were used for each model.

Model comparison using AIC supported the interaction model (main effect vs. interaction: 20282 vs. 20280). An evaluation of the β estimates using Satterhwaite's approximation for the interaction model revealed that cubic human likeness cross orientation was significant ($F(1,2236.1) = 7.0$, $p = 0.008$). The linear human likeness ($F(1,2247.4) = 170.9$, $p < 0.001$) and the interaction between the linear human likeness and the orientation ($F(1,2238.8) = 5.7$,

$p = 0.017$) were significant.

We conducted follow-up analyses for each orientation condition using the simple main model, which included the linear, quadratic, and cubic functions of the human likeness as independent variables. For the upright condition, the linear ($F(1,1090.4) = 255.2$, $p < 0.001$) and cubic ($F(1, 1100.0) = 4.6$, $p = 0.033$) functions of the human likeness were significant. For the inverted condition, only the linear human likeness was significant ($F(1,1114.2) = 162.4$, $p < 0.001$); the cubic function did not reach significance ($F(1, 1124.1) = 0.6$, $p = 0.459$).

In addition, the AIC-based model comparisons between the cubic and linear models for each orientation supported the cubic model for the upright condition (linear vs. cubic: 10110 vs. 10109) and the linear model for the inverted condition (linear vs. cubic: 10286 vs. 10290).

In summary, the results support the notion that 1) the relationships between the human likeness and the uncanniness ratings for the human and humanlike agents' expressions are cubic, and that 2) this processing depends on configural processing.

Differences between conditions

A within-participant ANOVA with actor type, orientation, emotion, and distortion as factors revealed significant interactions between type and distortion ($F(2,44) = 5.09$, $p = ,01$ $\eta^2_p = .19$), orientation and emotion ($F(2,44) = 19.28$, $p < .001$, $\eta^2_p = .3$), type and emotion ($F(2,44) = 3,46$ $p = .04$, $\eta^2_p = .14$), and all factors combined ($F(2,44) = 4.16$, $p = .022$, $\eta^2_p = .16$).

Post-hoc Tukey tests were conducted to test for differences between distortion levels across orientations and actor types. Results are presented as significant increases in uncanniness from one level of distortion to the next. For example, *0 to 2* refers to an increase of

uncanniness from distortion level 0 (base face) to level 2 (500ms delay). Data are summarized in *Figure 11.4* for human expressions, *Figure 11.5* for android expressions, *Figure 11.6* for CG expressions, and test statistics are summarized in *Tables 11.1* and *11.2*. In summary, for angry expressions, asynchrony increased uncanniness for upright, but not inverted human faces, it increased uncanniness for both upright and inverted android faces, and did not increase uncanniness for either upright or inverted CG faces. For happy expressions, asynchrony increased uncanniness for upright and inverted human faces, for upright (but not inverted) android faces, and for both upright and inverted CG faces (more so for inverted).

Table 11.1

Test statistics of each performed post-hoc tests of distortion (asynchrony) differences for across orientation, actor type, for angry expressions.

Emotion	Actor	Orientation	Distortion difference	t-value	p_{adj}-value	Effect size (d)
			0-1	$t(1060) = -1.36$	.26	
		upright	0-2	$t(1060) = -2.09$	.055	
			1-2	$t(1060) = -3.15$	.002*	0.63
	human		0-1	$t(1060) = 0.35$	1	
		inverted	0-2	$t(1060) = -0.88$	.57	
			1-2	$t(1060) = -1.46$	.22	
			0-1	$t(1060) = -2.26$	1	

		upright	0-2	$t(1060) = -3.86$	< .001*	0.78
			1-2	$t(1060) = -4.54$	< .001*	0.73
angry	android		0-1	$t(1060) = -2.32$	.031*	0.25
		inverted	0-2	$t(1060) = -3.64$	.003*	0.38
			1-2	$t(1060) = -0.98$	.485	
			0-1	$t(1060) = 0.51$	1	
		upright	0-2	$t(1060) = -0.92$	.54	
			1-2	$t(1060) = -1.77$	.116	
	CG		0-1	$t(1060) = 0.36$	1	
		inverted	0-2	$t(1060) = 0.04$	1	
			1-2	$t(1060) = 0.4$	1	

Table 11.2

Test statistics of each performed post-hoc tests of distortion (asynchrony) differences for across orientation, actor type, for angry expressions.

Emotion	Actor	Orientation	Distortion difference	t-value	p_{adj}-value	Effect size (d)
			0-1	$t(1075) = -3.2$	.002*	0.64
		upright	0-2	$t(1075) = -2.31$	.032*	0.46

			1-2	$t(1075) = -1.09$	1	
	human		0-1	$t(1075) = -3.93$	< .001*	0.73
		inverted	0-2	$t(1075) = -2.44$	.02*	0.46
			1-2	$t(1075) = -1.46$	1	
			0-1	$t(1075) = -4.24$	< .001*	0.87
		upright	0-2	$t(1075) = -3.16$	.002*	0.67
			1-2	$t(1075) = 1.23$	1	
happy	android		0-1	$t(1075) = -1.49$	.21	
		inverted	0-2	$t(1075) = -1.17$	.367	
			1-2	$t(1075) = 0.36$	1	
			0-1	$t(1075) = -2.75$	.009*	0.55
		upright	0-2	$t(1075) = -2.26$	.04*	0.45
			1-2	$t(1075) = 0.6$	1	
	CG		0-1	$t(1075) = -1.61$	.162	
		inverted	0-2	$t(1075) = -3.74$	< .001*	0.79
			1-2	$t(1075) = -2.72$	.01*	0.46

Figure 11.4

Mean uncanniness ratings for human expressions divided by emotion (angry, happy),,

distortion (asynchrony) level, and orientation. Error bars indicate standard errors. Asterisks indicate significant differences while NS indicate non-significant differences. Blue (first) significant marks are for upright, red (last) significant marks are for inverted conditions. For each emotion, differences were tested between distortion (asynchrony) levels 0 to 2 (upper line), 0 to 1 (lower left line), and 1 to 2 (lower right line), color-coded for orientation.

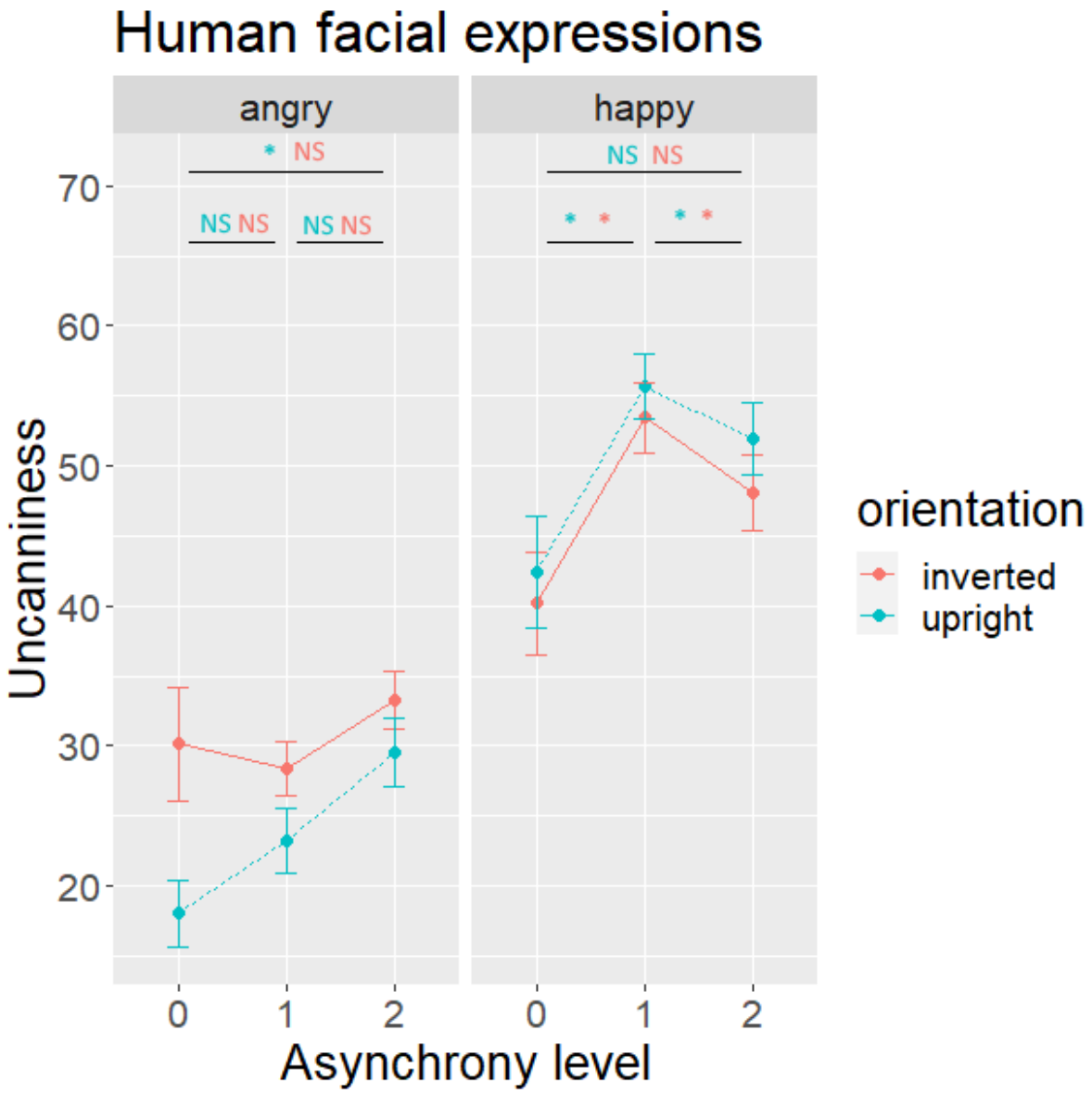


Figure 11.5

Mean uncanniness ratings for android expressions divided by emotion (angry, happy), distortion (asynchrony) level, and orientation. Error bars indicate standard errors. Asterisks

indicate significant differences while NS indicate non-significant differences. Blue (first) significant marks are for upright, red (last) significant marks are for inverted conditions. For each emotion, differences were tested between distortion (asynchrony) levels 0 to 2 (upper line), 0 to 1 (lower left line), and 1 to 2 (lower right line), color-coded for orientation.

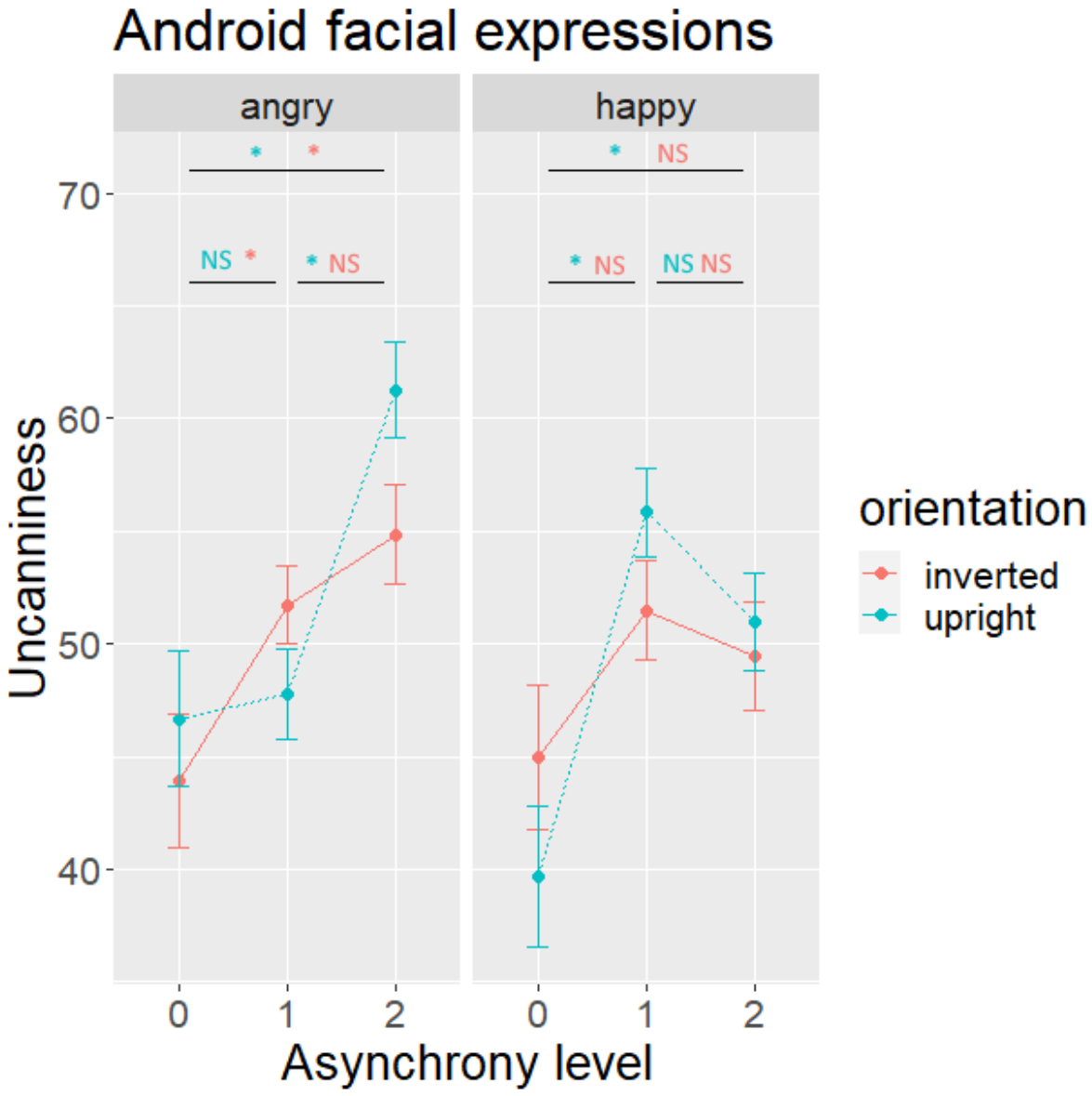
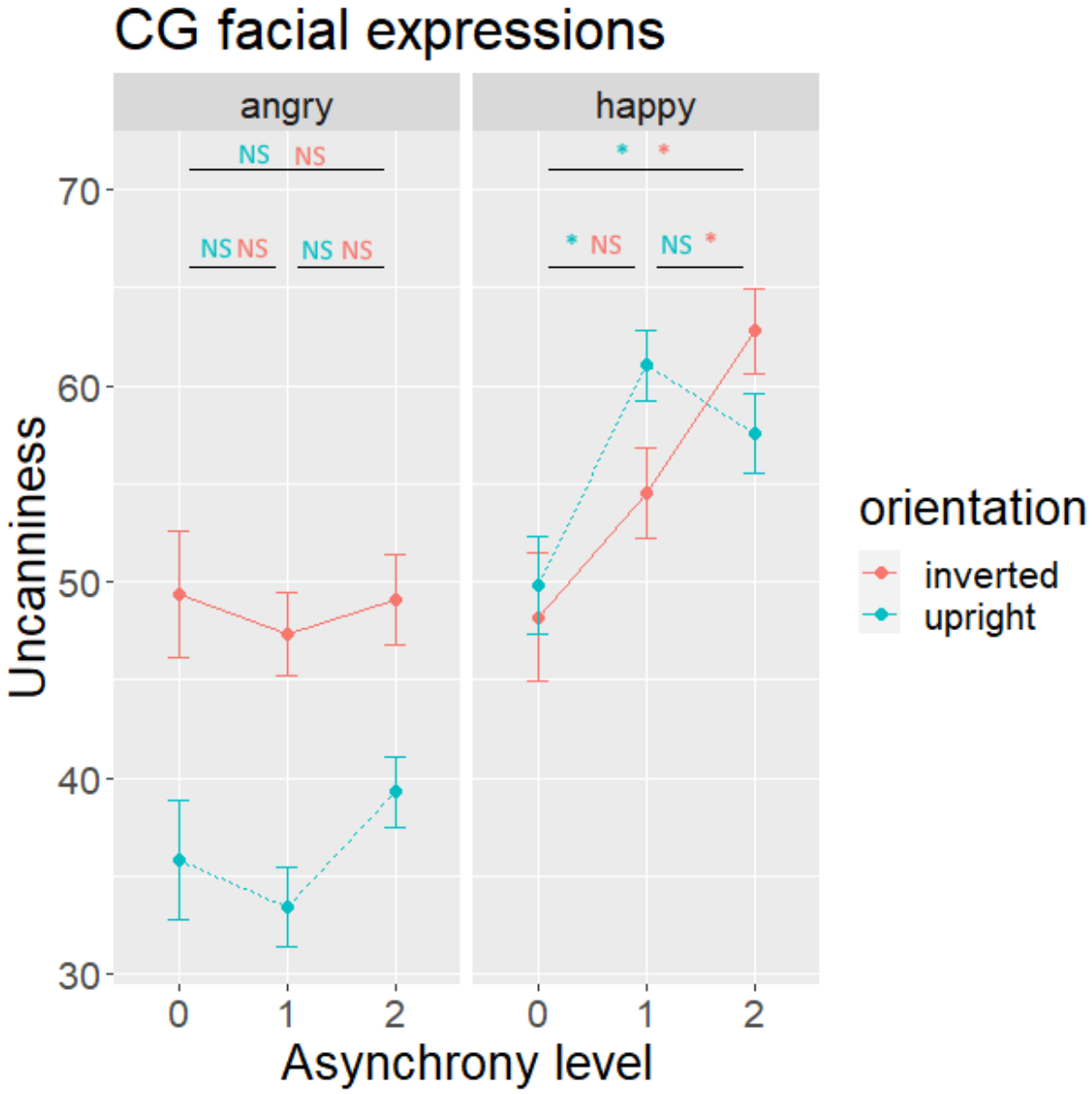


Figure 11.6

Mean uncanniness ratings for CG expressions divided by emotion (angry, happy), distortion (asynchrony) level, and orientation. Error bars indicate standard errors. Asterisks indicate significant differences while NS indicate non-significant differences. Blue (first) significant

marks are for upright, red (last) significant marks are for inverted conditions. For each emotion, differences were tested between distortion (asynchrony) levels 0 to 2 (upper line), 0 to 1 (lower left line), and 1 to 2 (lower right line), color-coded for orientation.



Discussion

Inversion effect and the polynomial uncanniness function

The present research investigated the effect of inversion, which is a proxy of configural

processing in faces and facial expressions, on the uncanniness of different agents' facial expressions across human likenesses. Differences between upright and inverted expressions were found. A cubic function of human likeness best explained the uncanniness of facial expressions, a result consistent with previous research (Diel et al., 2022; Mara et al., 2022). Only the linear function of human likeness was significant for the inverted facial expressions. Thus, a characteristic cubic function of human likeness for aesthetic appeal is only present when the configural processing of the facial expressions remains intact. This suggests that the typical observations on the relationship between the artificial agents' aesthetic ratings and human likeness depend on specialized processing mechanisms.

Although sigma-shaped relationships between human likeness and likability were found in a previous meta-analysis (Mara et al., 2022), they included work that lacked complete human stimuli. For research with a wider range of stimuli with various degrees of human likeness, cubic functions akin to Mori's (2012) uncanny valley are expected (Diel et al., 2022). Thus, the exact cause of the sigma-shaped uncanniness function remains unclear. Kätsyri et al. (2019), who also identified an "uncanny slope," suggested that this nonlinear relationship may be the result of a higher sensitivity to deviations in more familiar face categories. Since inversion reduces this sensitivity (Chapter 2), this "uncanny slope" was not found for inverted expressions in the present study.

We also found logistic patterns akin to two levels connected by an increasing slope in categorization tasks plotted against human likeness (Cheetham et al., 2011; Looser & Wheatley, 2010; MacDorman & Chattopadhyay, 2016). Categorization as human or non-human may consequently determine the effects of the ratings. Perhaps the disruption of this logistic function due to inversion indicates that categorization as human depends on configural processing. Alternatively, categorization and deviation-based aesthetic devaluation may both depend on configural processing, rather than categorization effects that cause

aesthetic devaluation.

Finally, perhaps the sigma-shaped function found here and in Kätsyri et al.'s research (2019) resulted from a limited stimulus range that excluded less humanlike stimuli like mechanical robot faces. A limited stimulus range has been suggested to be one cause of failing to find a "proper" uncanny valley function (Diel et al., 2022). Less humanlike (e.g., clearly mechanical) faces may have been evaluated as less uncanny, thus completing a *U*- or *N*-shaped cubic function (Mori et al., 2012) that is missing from the present data.

In any case, the results show that configural processing moderates the effect of human likeness on uncanniness. Specialized processing may act as a gateway to enhance the detection of errors or deviations that may produce negative evaluation results (Chapter 2; Chattopadyhay & MacDorman, 2017). Accordingly, the ratings of facial aesthetics are more sensitive when faces are presented upright instead of inverted (Bäuml, 1994; Leder et al., 2017; Santos & Young, 2008). The present results show for the first time that this role of configural processing on aesthetic appeal extends beyond facial structure and can also be applied to the dynamic expressions of faces. Furthermore, the results support the notion that specialized processing plays a role in evaluating artificial entities (e.g., Diel & MacDorman, 2021) and extends to the processing of dynamic facial expressions. Finally, the current results are in harmony with Mori's (2012) initial proposal that motion may enhance the uncanny valley effect. Temporal patterns of motions may increase uncanniness by providing additional dimensions from which a stimulus' appearance or behavior may deviate from the typical or expected patterns.

Asynchrony effects on dynamic expression uncanniness

In human expressions, asynchrony increased uncanniness in all but inverted angry faces. Furthermore, while asynchrony consistently increased uncanniness in upright human

expressions, it did so only for inverted happy (not angry) expressions.

In android expressions, a similar (yet opposite) pattern was observed: Asynchrony increased uncanniness in all but inverted happy faces, and asynchrony increased consistently for upright android expressions, but only for inverted angry (not happy) expressions.

Finally, in CG expressions, asynchrony effects on uncanniness were found in upright and inverted happy faces, but not upright and inverted angry faces. Finally, asynchrony increased uncanniness ratings (or did not increase those) regardless of expression orientation.

In summary, asynchrony tended to (inconsistently) increase uncanniness ratings across conditions. The effect however also depended on actor type and orientation: If changes in uncanniness increases through inversion are used as an indicator for configural processing, then the role of configural processing was observed for happy human and angry android expressions. Configural processing meanwhile seemed irrelevant for CG expressions.

Confirmations of the hypotheses give insight into the processing style of asynchronous motions in dynamic expressions: As the inversion effect is used as an indicator for configural processing, the results support the notion that deviations in an actor's facial expressions are detected using a configural processing style, and that this deviation sensitivity is more pronounced in android or human faces compared to CG faces.

Previous research has shown that the uncanniness of uncanny faces is reduced when faces are presented inverted, possibly due to a disruption of configural processing which is used to assess subtle distortions in a face (Chapter 2). Dynamic facial expressions may be processed by binding the sequence of face AU motions into a configural pattern. Deviations from this pattern, for example unusual timings of face AU motions in relation to the other units, may create an atypical expression, which is detected through the configural processing of the

expression dynamic and consequently negatively evaluated.

Previous research on static faces has shown that human faces increase the recruitment of face-specialized configural processing compared to CG faces (Miller et al., 2023). Similarly, humans are more sensitive to deviations in more realistic faces (Chapter 2). A higher level of realism of an actor may increase the sensitivity to deviations due to increased configural processing. However, the present research does not clearly support this view: There was no clear effect of actor on the effect of orientation on the uncanniness of asynchronous expressions, as CG and human faces were affected similarly.

Research on android and robot stimuli find an inversion effect which is smaller compared to human counterpart stimuli (Sacino et al., 2022; Schroeder et al., 2021; Zlotowski & Bartneck, 2013). However, the stimuli in these studies range from depictions of mechanical robots to close to humanlike androids, and the moderating role of actor realism on the inversion effect remains unclear. Our study meanwhile supports the notion that the uncanniness of android and human faces is at least partially processed configurally, which is not the case for CG faces.

In summary, the previous chapters established a new, refined theory of the uncanny, applied it to various categories, critically investigated its neural processes and against other theories, and showed that it can explain the uncanniness of a realistic androids. A detailed summary of the research and final evaluation of the refined theory follows in Chapter 12.

Chapter 12: General Discussion

Summary of results

The goal of this dissertation was to develop and test a refined theory of the uncanny valley focusing on a moderated linear function between familiarity or specialization and deviation/typicality on likability or uncanniness. Chapters 2 to 4 present research on the statistical relevance of a moderating familiarity or specialization variable; Chapters 5 to 6 focus on research applying the theoretical principles of the previous chapters onto inanimate categories; Chapters 7 to 10 rely on research critically investigating different theories on the uncanny valley with a focus on the refined theory; Finally, Chapter 11 presents research applying the refined theory to a real android. The research and results are summarized in more detail below.

Familiarity or specialization as a moderating variable

The uncanny valley is a nonlinear relationship between artificial entities' human likeness and likability: Near humanlike entities like androids are proposed to appear cold, eerie, or strange (Mori, 2012). Here it was investigated whether the uncanny valley stems from a higher sensitivity to atypical or distorted features caused by a higher level of perceptual specialization to certain stimulus categories, including human-related stimuli such as faces. Chapters 2 to 4 tested the validity of familiarity or specialization as a moderating variable on the effect of deviation on uncanniness.

Familiarity, orientation, and realism in faces. Human process conspecific faces in a specialized manner that takes configural information into account (Mondloch et al., 2002). Global inversion in face stimuli reduces specialization, and the level of difference in recognition ability of upright versus inverted faces (*inversion effect*) can be used as a marker of specialization (Farah et al., 1995). Less realistic faces (e.g., CG, robot, or schematic faces) show reduced inversion effects (e.g., Di Natale et al., 2023). Coincidentally, less realistic

faces also show a lower sensitivity to distortions (MacDorman et al., 2009; Mäkäräinen et al., 2014). If specialization increases the sensitivity to deviations, participants should have a higher tolerance for deviations in unfamiliar, less realistic, and inverted faces.

The experiment in Chapter 2 showed that in face stimuli, subjective familiarity as well as orientation and realism level (both correlating with level of specialization in faces), increased the sensitivity to detect subtle changes in faces, which in turn increased the uncanniness of distorted faces. Furthermore, it was found that a moderated linear function of face realism and distortion explained uncanniness better than a conventional quadratic "uncanny valley" function of human likeness, indicating a more accurate statistical model. In sum, the results support the notion that the level of familiarity or specialization increase the sensitivity to changes which then are judged as uncanny.

Specialization as a moderator variable. Chapter 2 showed that orientation and realism level, which have been associated with the level specialization, affected the sensitivity to uncanniness effects caused by face distortions. However, the link between realism and specialization level was presumed based on previous research, and not taken into account. Thus, a direct statistical link between level of specialization and uncanniness sensitivity is yet missing. Chapter 3 investigated whether level of specialization, measured via the face inversion effect, could predict the sensitivity to uncanniness caused by facial distortions. It was found that specialization predicted uncanniness changes caused by distortion, and that a moderated linear function of specialization and distortion could explain uncanniness better than a conventional uncanny valley plot: a quadratic function of human likeness. Thus, Chapter 3 provides a direct statistical link between the level of specialization and the sensitivity to deviation, and further supports the notion that such a moderated function is a more accurate statistical explanation of uncanniness effects than a conventional uncanny valley plot.

The causal role of specialization. Although Chapters 2 and 3 showed a statistical links between the level of familiarity or specialization on uncanniness sensitivity, the associations were merely correlational. Although a causal link between specialization and uncanniness sensitivity is theoretically plausible (specialization may increase attention to deviating details needed for evaluation at it does to discriminating ones needed for recognition), it remains to be empirically tested.

Chapter 4 investigated the causal link between specialization and uncanniness sensitivity by applying an expertise training paradigm (Gauthier et al., 1998): Expertise training consists of conducting various recognition tasks over the course 5 days to differentiate exemplars on a subordinate level using previously unknown *greeble* stimuli. Trained expertise shows behavioural (inversion effect; Gauthier et al., 1998) and neural (fusiform gyrus activity; Tarr & Gauthier, 2000) markers of perceptual specialization. Chapter 4 showed that after expertise training, distorted greebles are perceived as more uncanny compared to normal greebles or distorted greebles without expertise training. Meanwhile, there was no difference in uncanniness ratings between distorted and normal greebles without expertise training. Thus, Chapter 4 provides first evidence on a causal link between specialization and uncanniness sensitivity.

In total, Chapters 2 to 4 provide complementary evidence on specialization as a moderating variable for distortion on uncanniness, and how such a model is statistically superior to a conventional uncanny valley plot. While Chapter 2 shows that different manipulations of familiarity or specialization influence uncanniness sensitivity, Chapter 3 provides a direct statistical link of specialization. Finally, Chapter 4 shows a causal link between specialization and uncanniness sensitivity. Thus, Chapters 2 to 4 present the empirical framework for the validity of a moderating function of specialization, distortion, and uncanniness.

The refined theory in inanimate categories

The refined theory validated in Chapters 2 to 4 has, so far, only been applied to face and greeble stimuli. However, its theoretical base allows generalizations onto other stimulus categories, especially categories for which the level of specialization is measurable. Furthermore, investigating an uncanny valley or uncanniness effects in inanimate categories supports domain-independent theories on the uncanny valley in general. Thus, Chapters 5 and 6 investigated uncanniness effects in word and place stimuli, respectively.

Holistic word processing, ambiguity, and uncanniness in written text. Although the refined theory of uncanniness is mainly proposed as an explanation for the uncanny valley which is relevant for humanlike stimuli, it can be applied to inanimate categories as well. More than that, testing the predictions of the refined theory in inanimate categories, especially those with measurable degrees of specialization, would support the validity of the theory. One such stimulus category are written words which are processed in a holistic manner analogous to faces (Martelli et al., 2005). Participants are more sensitive to configural information in words of familiar languages (Wong et al., 2019). If the moderating effect of specialization on uncanniness sensitivity were correct, then configural orthographic distortions of words would be more uncanny in familiar compared to less familiar languages. Chapter 5 investigated this, and found that configural distortions in words increased uncanniness more in sentences written in a familiar compared to an unfamiliar language. Meanwhile, non-configural distortions affected uncanniness independent of language. Thus, Chapter 5 provided evidence that the specialization-dependent effects observed in Chapters 2-4 can also be applied to inanimate categories that show specialized processing, namely written text.

Liminal spaces and an uncanny valley of physical places. Although Chapter 5 found a moderating effect of specialization, orthographic distortion, and uncanniness in written words, it remains unclear whether this effect can be extended onto an uncanny valley

function. Chapter 6 aimed to investigate whether manipulations of realism (an alternative to a human likeness measure) lead to a non-linear, U- or N-shaped function of uncanniness in an inanimate object category, namely physical places. Furthermore, it was investigated whether this uncanny valley of physical places is driven by structural deviations akin to those observed for uncanny distorted faces, greebles, and words. As some naturally occurring physical places can appear eerie or uncanny (e.g., the internet phenomenon of *liminal spaces*; Wikimedia, 2023), it was investigated whether an uncanny valley of physical place could emerge using such stimuli. Indeed, an *N*-shaped function akin to an uncanny valley of realism could explain uncanniness ratings of artificial, real, and "liminal" places. As participants reported distorted features as a main cause of uncanniness, it was further investigated whether manipulations of the configural structure of physical places can increase uncanniness. It was found that structural deviations (lack, repetition, mismatch, or size changes of features like doors and furniture) increase uncanniness in physical places. Finally, absence or presence of people changed uncanniness ratings depending on whether people would be expected (e.g., in public places) or not (e.g., in private places). These results suggest that uncanniness in physical places is caused by whether the arrangement of features is consistent with place-depending expectations.

In sum, Chapters 5 and 6 extended the refined theory onto inanimate categories. Chapter 5 found that specialization sensitizes the effect of orthographic configural distortions (but not non-configural distortions) in words. Chapter 6 found that an uncanny valley in physical places can emerge, and that it is driven by structural deviations. In total, Chapters 5 and 6 provide complementary evidence that principles of the uncanny valley can be applied to inanimate objects, and that these effects may be explained by a moderating variable of specialization.

Critical investigation of multiple uncanny valley theories

Uncanny valley research is marked by a variety of, oftentimes competing, theoretical explanations (Diel & MacDorman, 2021; Wang et al., 2015). Chapters 7 to 10 serve to present research critically investigating theories by creating experimental designs testing multiple theories simultaneously.

The uncanny valley in voices. Although an uncanny valley in voice stimuli has been speculated on in the past, previous research consistently failed to find an uncanny valley of voices (e.g., Kühne et al., 2020). This may have been due to a limited range of stimuli: If current TTS voices are capable of avoiding an uncanny valley, only using TTS and real voices may be insufficient to generate a vocal uncanny valley even if some voices may actually be uncanny. Chapter 7 presents research investigating the existence of a vocal uncanny valley using distorted and naturally pathological voices in addition to TTS and real voices. In addition, it is investigated whether categorical ambiguity, a common theory of the uncanny valley (Cheetham et al, 2014) drives voice uncanniness. It was found that uncanny-valley like shapes do emerge when distorted and pathological voices are added, but the relationship between human likeness and uncanniness is linear when only TTS and real voices are used. Furthermore, categorical ambiguity could not explain voice uncanniness, and some unambiguous voices were uncanny (e.g., pathological voices) while the most ambiguous voice (a TTS voice) was not uncanny. A second study investigated whether uncanny voices are so because they have mind or animacy attributed to them, another explanation of the uncanny valley (Gray & Wegner, 2012). Results did not support that voices were uncanny because they had human-like qualities attributed to them. However, it was found that perceived “organic-ness” of a voice increased the sensitivity to uncanniness. Thus, distortions seemed to cause uncanniness especially in voices perceived as organic;

these results are consistent with disease avoidance accounts of the uncanny valley (MacDorman & Ishiguro, 2006).

In sum, the results suggest that an uncanny valley can emerge for auditory stimuli only, specifically human voices. Distorted and pathological voices were perceived as uncanny, although not because of mind/animacy misattribution or categorical ambiguity. Instead, voices may be uncanny when they are perceived as organic yet deviating, which would be expected from disease avoidance mechanisms. Finally, TTS voices successfully manage to escape an uncanny valley of voices.

Effect of uncanniness priming. Emotional priming is an effective tool to investigate effects of specifical emotional reactions on stimulus processing. The uncanny valley has been linked to disease avoidance mechanisms (MacDorman & Entezari, 2015) and mortality salience (MacDorman, 2005) in past research. If the uncanny valley were linked with either of those, then uncanny primes should have effects analogous to disgust or fear primes respectively. Chapter 8 investigated the effect of uncanny primes (compared to control, disgust, and fear primes) on reaction times towards disease- and death-related words in a lexical decision task. Disgust primes either increased or decreased reaction times towards disease-related words, while fear primes increased reaction times towards death-related words. Uncanny primes (consisting of showing participants images of uncanny androids) meanwhile did not affect disease and death word processing, analogous to control primes. Results show that uncanny primes do not work the same as disgust or fear primes, which is evidence against disease avoidance or mortality salience explanations of the uncanny valley.

In additional tasks, it was found that body distortions increased uncanniness, but that this effect was reduced when inverted – supporting the role of configural processing in the uncanny valley. Furthermore, inversion decreased uncanniness in some, but not all

ecologically valid uncanny stimuli (videos of uncanny androids or CG characters). These results indicate that configural processing plays a role in some uncanny stimuli, but that other effects or mechanisms may also contribute to stimulus uncanniness.

Electrophysiological correlates of uncanny faces. Neurophysiological correlates to behavioural tasks can add a deeper understanding in the mechanisms underlying specific types of processing. The refined theory suggests that the uncanny valley underlies increased processing activity in specialized categories, which is in line with processing disfluency theories (Winkielman et al., 2003) and predictive coding (Friston, 2010) accounts. For faces, increased activity would be expected in face-related markers, such as the P100 component associated to face feature processing and the N170 component related to face configuration processing (Olivares et al., 2015). Chapter 9 presents behavioural and neurophysiological research linking the uncanny valley to face-related processing. A behavioural task found that an uncanny valley is reduced when faces are presented inverted instead of upright, supporting a configural processing account of the uncanny valley. The neurophysiological task showed that upright (not inverted) distorted uncanny faces elicited higher N170 amplitudes while distorted faces elicited higher P100 amplitudes regardless of orientation. As N170 amplitudes are markers of configural processing and P100 amplitudes markers of feature processing, research may show increased processing need for distorted configural faces when presented upright (N170), and distorted features that are not influenced by orientation (P100). In sum, the results support the notion that the uncanny valley is related to increased domain-specific and specialized processing.

Individual differences sensitizing uncanniness across domains. The effect of individual differences on uncanniness ratings can give insight into which specific cognitive processes are related to evaluations of uncanniness (MacDorman & Entezari, 2015). Five candidate differences are deviancy aversion (Gollwitzer et al., 2017), disgust sensitivity (MacDorman

& Entezari, 2015), need for structure (Lischetzke et al., 2017), anxiety-neuroticism (Goldberg et al., 1999; MacDoramn & Entezari, 2015), coulrophobia (Tyson et al., 2023). Chapter 9 presents research investigating the effect of these individual difference measures on the uncanny valley, uncanniness of clowns, and uncanniness effects across different stimulus domains. Results show that individual difference variables predicted different uncanniness effects: While disgust sensitivity and anxiety predicted android uncanniness akin to previous research (MacDorman & Entezari, 2015), deviancy aversion predicted domain-independent uncanniness effects, while coulrophobia predicted uncanniness of distorted human features (faces, voices, bodies). Furthermore, the results emphasize a potential link between the uncanny valley and fear of clowns. Taken together, the results support the notion of the heterogeneity of the uncanny valley: Multiple mechanisms like deviancy aversion or disease avoidance underlie uncanniness which get cumulated on the uncanniness of androids.

The refined theory applied to realistic dynamic android expressions

While the previous chapters focused on validating and testing the refined theory in different stimulus categories and against other theories, only Chapters 8 and 10 investigated the relevance of the refined theory in ecologically valid stimuli, namely uncanny android and CG characters. Chapter 8 found that inversion decreases uncanniness in some uncanny stimuli, while Chapter 10 showed that deviancy aversion sensitizes the uncanniness of uncanny static android images.

To further investigate the relevance of the refined theory for androids, Chapter 11 (Diel, Sato, Hsu, & Minato, in review) presents research using dynamic face emotion expressions of CG faces, humans, and the android Nikola. Nikola has been developed to imitate realistic human expressions (Sato et al., 2022). Results of Chapter 11 show that while a cubic function of human likeness and uncanniness emerged for upright expressions, the function was reduced to a linear one when inverted. Furthermore, while asynchronous expressions mostly increased

uncanniness, inversion tended to reduce this effect in android and human faces, but not in CG faces. As CG faces are marked with lower inversion effects (Crookes et al., 2015), results can be explained by specialized processing in dynamic expression sensitizing the uncanniness of asynchrony (which can in turn be understood as configural deviations). Thus, the results support the notion that the refined theory can be applied to realistic androids and, more specifically, to the processing of dynamic facial expressions.

Summary

The goal of the dissertation was to test a refined theory of the uncanny proposing a moderated linear function between specialization, deviation, and uncanniness. Chapter 2 to 4 provided evidence of the statistical, causal association between specialization and the sensitivity to configural distortions. Chapters 5 and 6 extended this effect onto inanimate categories. Chapters 7 to 10 tested multiple theories of the uncanny valley, again finding evidence in favour of the refined theory. Finally, Chapter 11 successfully applied the refined theory onto realistic, ecologically relevant androids. Thus, the dissertation provides substantial evidence of the refined theory as a cause of uncanniness and its relevance to the uncanny valley. However, research (notably Chapters 7 and 10) also found relevance for other potential processing mechanisms for the uncanny valley, such as disease avoidance. Thus, while the refined theory can explain uncanniness across categories, the uncanny valley may still emerge through a multitude of mechanisms (Bartneck et al., 2007; Diel & MacDorman, 2021) and may also depend on the exact type of stimulus or manipulation used. However, a moderating role of specialization on distortion effects on uncanniness as a contributor to the uncanny valley can, given the presented research, not be ignored.

Evaluation of the refined theory

The refined theory of the uncanny presented here provides several advantages over the conventional uncanny valley model (Mori, 2012). Especially as a statistical model the refined

moderator model outweighs a conventional cubic uncanny valley with empirical, statistical, and theoretical benefits which are discussed below.

Advantages of new theory

Statistical plausibility and accuracy. Science prioritizes the simplest explanations for the widest range of observations. The contemporary uncanny valley function suggests a specific N-shaped cubic relationship between human likeness and uncanniness. This model suffers in parsimony due to its complexity and because such cubic statistical functions are rare in nature. By adding a third variable, a moderated linear function provides a statistically simpler and more plausible explanation to uncanniness effects than the uncanny valley. Furthermore, multiple studies presented here (notably in Chapters 2, 3, and 7) found that moderated linear functions were more accurate statistical fits than the uncanny valley model.

Thus, the moderating model provides a simpler yet more accurate representation of the data than the uncanny valley.

Generality. Not only is the moderated linear function a simpler and more accurate statistical fit compared to the uncanny valley when given the same observations; the moderated linear function can also explain a wider range of observations than the uncanny valley model. For example, although uncanniness effects were observed for inanimate objects (see Chapters 5 and 6) the uncanny valley model with its focus on a human likeness axis did not provide satisfying explanations on those. The uncanny valley also does not explain moderating effects of face realism and distortion on likability observed in past research (MacDorman et al., 2009; Mäkäräinen et al., 2014), while the refined theory is capable of putting these observations into a consistent statistical pattern (Chapters 3 and 4).

Furthermore, the refined theory can be extended to deviations or anomalies that have been left untouched in uncanny valley research, such as those observed in people with physical

disabilities or deformities: Chapter 7 found that pathological voices appear uncanny, and Chapter 9 showed that uncanny faces and faces with natural disfigurements express similar behavioural and neurophysiological responses (see also Diel & MacDorman, 2021). Thus, the refined theory is capable of synergizing research on the uncanny valley and on the negative evaluation of naturally non-typical faces (Stone, 2021; Workman et al., 2021).

In summary, the refined theory can be applied to a broader range of observations compared to the uncanny valley, and can portray these observations in a more satisfying statistical pattern.

Accuracy of predictions. The uncanny valley has been criticized for its variables' lack of clear definitions: For example, human likeness is a multidimensional construct that can be measured or manipulated in multiple ways, including objective and subjective human likeness estimates (Burleigh et al., 2013; Diel et al., 2022; Ho & MacDorman, 2017). Furthermore, predictions become confusing when non-human stimulus dimensions like animal stimuli are chosen, and would then need to be replaced with realism or zoomorphism (Schwind et al., 2018). Finally, when research fails to find an uncanny valley (e.g., Bartneck et al., 2009; Cheetham et al., 2014; Cheetham et al., 2015; Kätsyri et al., 2019), it is unclear whether such results should be considered evidence against the uncanny valley or whether they emerged due to an inadequate operationalization of human likeness (e.g., using morphing stimuli with an inadequate range of human likeness to capture the complete cubic function). The refined theory of the uncanny meanwhile is marked by clearly defined variables and easy empirical testability: As long as a stimulus' degree of specialization is measurable (e.g., inversion effect in faces) or can at least be presumed (e.g., language familiarity in written text), then the interaction between specialization and incremental levels of distortion on uncanniness can be tested across multiple stimulus domains.

Results that contradict the uncanny valley. Although sparsely, some past studies failed to find an uncanny valley (Bartneck et al., 2009; Cheetham et al., 2014; Cheetham et al., 2015; Kätsyri et al., 2019). While the researchers suggested their results to contradict the uncanny valley model, the refined theory proved an explanation on why such results may have occurred.

Most research studies finding no evidence of an uncanny valley used a range of face stimuli with similar facial proportions. If said proportions are acceptable in a real face, the refined theory presented in this dissertation would not predict morphed version of these faces to be uncanny (but to be more sensitive to distortions *if* any are present). First, Cheetham et al. (2014) found a "happy valley" in which the most ambiguous faces morphed between human and avatar faces were rated more positively; However, it was presumed that valence should be negatively associated with perceptual discrimination difficulty (high discrimination difficulty was equated to the degree of human likeness at which the uncanny valley occurs), which was not postulated in Mori's (2012) proposal and later discredited (e.g., Mathur et al., 2020). Thus, perceptual discrimination or categorization difficulty should not be automatically equated with the occurrence of the uncanny valley. Furthermore, Morphing avatar and human faces with analogous facial proportion would create intermediate stimuli without distortions which would avoid an uncanny valley despite being categorically ambiguous. A similar argument can be made for Cheetham et al. (2015) who found no evidence of an uncanny valley on a continuum of face morphs. Analogously, Kätsyri et al. (2019) found an "uncanny slope" rather than an uncanny valley when using faces in different levels of realism. Again, faced did not differ in their facial proportions given that the low-realism renderings were based on the original real human face.

The refined theory of the uncanny can contextualize these results as a range of faces differing on the level of realism but without any noteworthy deviations: The refined model does not

exclude the existence of intermediately humanlike faces that are not uncanny; rather, it proposes that the reason *why* so many uncanny stimuli are near humanlike is because potential errors become more apparent due to specialized processing. If intermediate humanlike faces however do not contain deviating facial proportions, for example because these faces were morphed with less realistic faces with the same proportions, then the refined model would not expect any uncanniness effects to occur. Consistent with this idea, morphing studies using more proportionally different faces as endpoint stimuli do find uncanny valley effects: For example, Yamada et al. (2013) morphed normal human faces with faces of the cartoon character Charlie Brown and found that intermediate morphs were less likable. Similarly, Seyama and Nagayama (2007) found that morphs between humans and dolls were more eerie. Given that cartoon and doll faces tend to have unrealistic facial proportions, the refined model would expect that by increasing their realism levels (by being morphed with a real human face), their exaggerated proportions would become more apparent due to specialized processing, eliciting uncanniness effects. Thus, the refined model would predict that morphing stimuli with similar facial proportions should not elicit uncanniness effects, whereas morphing human faces with faces containing exaggerated proportions (e.g., stylized cartoon, robot, or doll faces) would lead to uncanniness effects; specifically, uncanniness effects should occur *because* specialized processing is increased by morphing with a real human face, which then sensitizes the processing of unrealistic proportions.

Finally, Bartneck et al. (2009) found that a realistic android was not liked less compared to a real human in a real-life interaction. However, a further study did show uncanniness effects using different models of same android (Geminoid HI), so these negative results are challenged by a lack of replication (Zlotowski et al., 2015).

Finally, Mori (2012) initially proposed that motion should increase the uncanny valley effect. However, this prediction has been considered disproven by several researchers (Kätsyri et al., 2015; Piwek, 2014). The refined theory meanwhile not only offers motion effects another chance but is also capable of providing explanations on why motion effects did not occur in previous studies.

As biological motion (e.g., in facial expressions) can be processed in a specialized manner, deviations in motion alone (without manipulating the static structure) can produce uncanniness effects (see Chapter 11). When both static and dynamic deviations co-occur, e.g., in an imperfectly designed android stimuli which also shows error in its motions, then accumulations of deviation-uncanniness effects across different modalities are expected. Thus, rather than proposing that "movement amplifies the uncanny valley effect" (see Kätsyri et al., 2015), the refined theory would expect that "dynamic deviations in biological motion increase uncanniness in addition to static deviations, thus amplifying an uncanny valley". As degrading motion quality has been found to decrease likability (Piwek et al., 2014; Thompson et al., 2011), adding both dynamic and static deviation can be expected to shoe cumulative effects. Meanwhile, consistent with previous research, merely animating inanimate characters would increase likability due to an increase in human likeness (McDonnell et al., 2012; Piwek et al., 2014) which has previously been interpreted as a falsification of the motion hypothesis (Kätsyri et al., 2015). Thus, the refined theory of the uncanny is able to give potential motion effects another chance (Mori, 2012), without going into conflict with previous research finding no motion effects. Such predicted motion effects may be the subject of future research.

Compatibility with neurocognitive theories. Finally, the refined model presented here is built on cognitive theories on specialized (face) processing and compatible with neurocognitive

models of stimulus evaluation, notably processing disfluency and expectation violation in predictive coding.

It has been long established that additional processing dimensions are recruited for stimulus categories for which individual-level recognition and discrimination are important, notably human face stimuli, including the recruitment of specialized brain areas like the FFA (Kanwishe, McDermott, & Chun, 1997). Several studies also show that these additional processing dimensions (e.g., configural processing) are used in the aesthetic evaluation of face stimuli (Bäuml, 1994; Leder, Goller, Forster, Schlageter, & Paul, 2017; Santos & Young, 2008). Increased neural activity in specialized areas has been observed for atypical stimuli belonging to the category: For example, distorted faces elicit stronger face-related neural activity (Carbon et al., 2003; Cassia, Kuefner, Westerlund, & Nelson, 2006; Hahn et al., 2012; Halit, de Haan, & Johnson, 2000; Löffler, Yourganov, Wilkinson, & Wilson, 2005; Mattavelli et al., 2012; Milivojeric et al., 2003; Rothstein et al., 2001; Said et al., 2010; Todorov et al., 2013; Workman et al., 2021). Analogously, in Chapter 9 I presented research that uncanny faces elicit increased face-related activity, specifically N170 amplitudes for configural deviations and P100 amplitudes for feature-level deviations.

Increased activity in specialized areas for unusual stimuli have also been observed for physical place stimuli (Rémy et al., 2013) and voice stimuli (Andics et al., 2010; Latinus et al., 2013). Analogous effects on body or written text processing have not yet been investigated. A domain-independent pattern of increased activity in category-specific areas may be related to the aesthetic devaluation of deviating stimuli described in this dissertation.

Such increased activity for atypical or distorted exemplars in a specialized brain area can occur for two reasons, one being stimulus processing disfluency and the other expectation violations in predictive coding. Stimuli distant from prototypical appearance increase

processing disfluency which in turn decreases aesthetic evaluation (Winkielman et al., 2003). Increased activity in specialized areas for atypical or distorted stimuli may reflect increased processing need (Olivares et al., 2015). Thus, the increased neural activity observed for uncanny stimuli may reflect disfluency caused by deviations. Alternatively, deviating stimuli can elicit errors in predictive coding (Friston, 2010; Keller & Flogel, 2018), and increased activity in specialized areas may correspond to such error signals, leading to an uncanny valley effect (Saygin et al., 2012; Urgen et al., 2018).

In any case, the refined theory of the uncanny is built on previous research in specialized processing while deviation-based effects are compatible with established neurocognitive theories on stimulus evaluations. Further research may focus on which exact neurocognitive processes underlie uncanniness effects that have been topic in this work.

Disadvantages of the refined model

Lack of ecologically valid uncanny stimuli. The research presented in this dissertation mainly relied on one type of stimulus manipulation: The (sometimes incremental) distortion of feature proportions. Effects of such distortions on uncanniness ratings have been consistently observed, including moderating effects of direct or indirect markers of specialization.

However, it remains unclear whether the uncanniness effects caused by distortions are the same that cause androids, CG characters, or other entities typically described to fall into the uncanny valley to appear uncanny. Only three chapters presented here (Chapters 8, 10, and 11) included ecologically valid android stimuli: Chapter 8 found that only some android stimuli show reduced uncanniness when inverted; Chapter 10 found that a quadratic human likeness function of uncanniness including uncanny androids is sensitized by individual differences in deviancy aversion; and Chapter 11 found that asynchronies in dynamic facial expressions in androids show inversion effects analogous to human stimuli. While these

results are promising, future research can establish a more direct role of specialized processing in the evaluation of uncanny android stimuli.

Not all uncanny stimuli show an inversion effect. Chapter 8 found that inversion reduced uncanniness in only some android stimuli. If inversion is a marker of specialized processing, then these results do not support its role in the uncanniness of androids in general. It is possible that certain specialized processing mechanisms survive inversion, or that deviations can cause uncanniness even on a feature-level (see also Chapter 9 on increased P100 activities in inverted distorted faces). Alternatively, results could support that other mechanisms unrelated to specialized processing contribute to uncanniness ratings.

Other causes of uncanniness. In the same vein as the previous point, the results presented in this dissertation do not negate the existence of other causes of uncanniness. Although some results find evidence against prevalent theories on the uncanny valley, such as ambiguity-processes (Chapters 5 and 7) or disease avoidance (Chapters 8 and 10), other mechanisms may still increase uncanniness especially when the role of configural processing is questions (e.g., the lack of inversion effect on uncanniness ratings in Chapter 8). For example, disease avoidance may still influence aesthetic ratings in stimuli other than those used in Chapters 8 (see also Ho et al., 2008; MacDorman & Entezari, 2015). As no individual difference variable has been consistently associated with uncanniness in Chapter 10, the mechanisms underlying uncanniness effects may be depend on the stimulus and manipulation.

Evidence for a mathematical specification. Although the refined model was introduced as a mathematical moderated linear function, the presented research have not directly tested the mathematical predictions. The research in Chapter 3 was closest to express a moderated linear function as introduced in Chapter 1 by finding that the level of face inversion effect (an indicator of face specialization) moderated the effect of deviation on face uncanniness within

a given face category. However, the exact beta values of the interaction terms were not investigated. It would be expected that for categories with a higher specialization, beta values for deviation on uncanniness would be higher, represented by steeper linear slopes.

Further research presented here has not directly investigated the proposed mathematical model. Instead, the effects of specialization and deviation on uncanniness were often tested via group differences (e.g., changes in uncanniness across deviation levels for stimuli with a high versus low level of specialization). While such results can indicate specialization effects of deviation (e.g., deviation effects on uncanniness are stronger for upright compared to inverted faces), they are not representations of the proposed mathematical model. Thus, if the refined theory is strictly considered as the proposed moderated linear function in a mathematical sense, the results would not sufficiently support it. Instead, the results support the notion that indicators of specialization interact with the effect of deviation on uncanniness in that such effects are higher in stimulus categories with higher specialization indicated by stimulus familiarity, orientation, realism, or expertise (see Chapters 2, 4, 5, 7, 8, 11).

As the mathematical model proposes specific predictions on beta values (i.e., higher values for deviation effects in categories with a higher level of specialization) that have not been tested in the present work, future research may aim to test such exact predictions by comparing relevant beta values.

Participant selection and generalizability. Participants in the presented research were mainly young adult psychology students from the UK, and in some cases participants from the UK, US, Germany, or Japan that were not exclusively young adults or psychology students. The focus of a young adult Western or Japanese populations may limit the generalizability of the results: For example, older participants may not experience an uncanny valley (Tu, Chien, & Yeh, 2020). Analogously, the devaluation of anomalous faces may be culturally moderated

(Workman et al., 2021). However, age effects of uncanniness may be due to age effects in face processing, as specialization for faces is stronger for faces of individuals in a similar age range (Lamont, Stewart-Williams, & Podd, 2005), and face specialization decreases with older age (Connolly, Young, & Lewis, 2021). In this sense, the selection of a limited age range for participants is beneficial for the investigation of specialized processing in faces of individuals that are also of a similar age, as was the case with most face stimuli used in this research. Meanwhile, a lack of uncanny valley for older participants (Tu et al., 2020) may have been observed because older participants were less specialized to the faces presented. Similarly, face specialization effects also depend on experience with different ethnicities (see other-race-effect; Rhodes et al., 2006, which may also play a role for the uncanny valley effect in faces, Saneyoshi et al., 2022). Thus, focusing on specific cultures for research using faces of common ethnicities (Caucasian and Asian faces for European and Japanese participants, respectively) improves control over face specialization effects that are critical for the current research.

Nevertheless, the limited demographical range of participants does not answer questions of generalizability. Future research may, for example, investigate whether an uncanny valley effect does occur for older participants specifically when faces of older individuals are used.

Measures and conceptualization of uncanniness. The present research was based on self-report scales of uncanniness and related concepts as established adequate measures of the UV effect (Diel et al., 2022). Although measures related to “uncanniness” and “strangeness” are both used in UV research (Ho & MacDorman, 2017), they may be considered different concepts (Diel et al., 2022) and accordingly, such items do not always highly correlate and may depend on stimulus category (e.g., see the intercorrelation for word stimuli in Chapter 11). If strangeness and uncanniness can be considered two different concepts, then arguably, they both play a role in the UV effect (Ho & MacDorman, 2017). However, if the goal is to

investigate uncanniness effects specifically, measures of strangeness may be inadequate especially in cases of lower intercorrelation.

Furthermore, it is unclear whether participants actually shared analogous experiences when they reported higher uncanniness ratings for different stimulus categories. Self-reports in uncanniness may, at least partially, represent changes in general affect caused by factors which differ across stimulus categories when stimuli are distorted. As no measure of uncanniness or the UV effect has been tested for discriminant validity, such confounding factors cannot be excluded, and the validity of self-report measures of uncanniness across different stimulus categories cannot be guaranteed.

Finally, the concept of "uncanniness" has not yet been properly established, and although multiple ideas have been proposed (Benjamin & Heine, 2023; Diel et al., 2022; Mangan, 2015), no consensus on its nature (e.g., whether it is an experience, an aesthetic, a sensation, a feeling, an emotion, etc.) or properties (e.g., potential cognitive and physiological changes) is present. Although self-report rating scales are considered effective and well-established measures (Diel et al., 2022), their exact relation to the subjective experience of uncanniness remains unclear.

However, these issues are prevalent in the UV research field in general, and as the field grows and develops proper measures and conceptualizations of uncanniness and related constructs, such developments can be used to more accurately assess whether the uncanniness effects observed in the present works represent the same concept.

No explanation of the uncanny feeling. It has been a long-standing criticism of domain-independent, cognitive theories of the uncanny valley that such theories do not offer any explanations on uncanniness as a specific sensation (e.g., MacDorman & Entezari, 2015). Yet several researchers noted that the uncanny valley effect is marked by a specific sensation or

experience of eeriness or uncanniness (Diel et al., 2022; MacDorman & Entezari, 2015; Mangan, 2015; see also Benjamin & Heine, 2023). General domain-independent neurocognitive mechanisms like processing disfluency and predictive coding may be linked to stimulus evaluation, they make no assumptions on uncanniness specifically. Furthermore, such processes can be observed over a wide range of stimuli and situations which are not typically described as "uncanny". Although MacDorman and Entezari (2015) proposed that uncanniness specifically may be linked to disease avoidance mechanisms, evidence for these mechanisms have not been found in this dissertation. Future research can aim to investigate in which circumstances mechanisms like disfluency or prediction errors cause uncanniness and in which it does not.

In summary, the present work is limited by the following points: 1) inconsistent results and lack of research using android or similar stimuli; 2), the potential of heterogeinity; 3) lack of testing of the refined model as a mathematical prediction, 4) generalizability of the results; 5) unclarity regarding the conceptual framework of one of its variables (uncanniness) which is a prevalent problem in the research field, and an explanation of the "uncanny feeling". Thus, while this work presents evidence that deviation sensitivity and its negative evaluation is higher in certain categories that also express higher specialization, it remains unclear whether this is the only relevant mechanism of uncanniness, whether it is the relevant mechanism for ecologically valid stimuli, whether the uncanniness measured is analogous to the uncanniness indicative of the UV effect, and whether the moderated linear function is mathematically sound to describe the effect.

Conclusion

In my work, I aimed to develop and test a neurocognitive model of uncanniness and the uncanny valley. Over the course of 17 experiments, the role of perceptual specialization on the sensitivity to deviations has been tested across multiple domains, including faces, bodies,

voices, places, written text, and biological motion. Furthermore, the model provides explanations on the devaluation of naturally deviating biological stimuli like disfigured faces or pathological voices which can occur with social stigma of individuals with disabilities. In parallel, multiple contemporary theories on the uncanny valley have been critically evaluated, like categorization ambiguity or disease avoidance. With my work I offer a statistically simpler and more accurate, general, and theoretically plausible model that sheds light into the causes of the uncanny feeling.

References

1. Abrams, D., Hogg, M. A., & Marques, J. M. (2005). A Social Psychological Framework for Understanding Social Inclusion and Exclusion. In D. Abrams, M. A. Hogg, & J. M. Marques (Eds.), *The social psychology of inclusion and exclusion* (pp. 1–23). Psychology Press
2. Abubshait, A., & Wiese, E. (2017). You look human, but act like a machine: Agent appearance and behavior modulate different aspects of human–robot interaction. *Frontiers in Psychology*, *8*, Article 1393. https://doi.org/10.3389/fpsyg.2017.01393
3. Adams, A. & Robinson, P. N. (2011). An Android Head for Social-Emotional Intervention for Children with Autism Spectrum Conditions. *Lecture Notes in Computer Science*, 183–190. https://doi.org/10.1007/978-3-642-24571-8_19

4. Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, *74*, 187–195. https://doi.org/10.1016/j.jesp.2017.09.004

5. Alkhaldi, R.S., Sheppard, E. & Mitchell, P. (2019). Is There a Link Between Autistic People Being Perceived Unfavorably and Having a Mind That Is Difficult to Read?. *Journal of Autism Development Disorder*, *49*, 3973–3982. https://doi.org/10.1007/s10803-019-04101-1

6. Almaraz, S.M. (2017). Uncanny processing: Mismatches between processing style and featural cues to humanity contribute to uncanny feelings. 43803401

7. Altenberg, E. P., & Ferrand, C. T. (2006). Fundamental frequency in monolingual English, bilingual English/Russian, and bilingual English/Cantonese young adult women. *Journal of voice : official journal of the Voice Foundation*, *20*(1), 89–96. https://doi.org/10.1016/j.jvoice.2005.01.005

8. Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: metacognitive difficulty activates analytic reasoning. Journal of experimental psychology. General, 136(4), 569–576. https://doi.org/10.1037/0096-3445.136.4.569

9. Alter, A.L., & Oppenheimer, D.M. (2008). Easy on the mind, easy on the wallet: The roles of familiarity and processing fluency in valuation judgments. *Psychonomic Bulletin & Review 15*, 985–990. https://doi.org/10.3758/PBR.15.5.985

10. Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the enigmatic face: the importance of facial dynamics in interpreting subtle facial expressions. *Psychological science*, *16*(5), 403–410. https://doi.org/10.1111/j.0956-7976.2005.01548.x

11. Amir, O., & Levine-Yundof, R. (2013). Listeners' attitude toward people with dysphonia. *Journal of voice : official journal of the Voice Foundation, 27*(4), . https://doi.org/10.1016/j.jvoice.2013.01.015

12. Amoruso, L., Sedeño, L., Huepe, D., Tomio, A., Kamienkowski, J., Hurtado, E., Cardona, J. F., Álvarez González, M. Á., Rieznik, A., Sigman, M., Manes, F., & Ibáñez, A. (2014). Time to Tango: expertise and contextual anticipation during action observation. *NeuroImage*, *98*, 366–385.

13. Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*(3), 409–429. https://doi.org/10.1037/0033-295X.98.3.409

14. Andics, A., McQueen, J. M., Petersson, K. M., Gál, V., Rudas, G., & Vidnyánszky, Z. (2010). Neural mechanisms for voice recognition. *NeuroImage, 52*(4), 1528–1540. https://doi.org/10.1016/j.neuroimage.2010.05.048

15. Appel, M., Izydorczyk, D., Weber, S., Mara, M., & Lischetzke, T. (2020). The uncanny of mind in a machine: Humanoid robots as tools, agents, and experiencers. *Computers in Human Behavior*, *102*, 274–286. https://doi.org/10.1016/j.chb.2019.07.031

16. Appel, M., Weber, S., Krause, S., & Mara, M. (2016). On the eeriness of service robots with emotional capabilities. 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Christchurch, New Zealand. 411 – 412. https://doi.org/10.1109/HRI.2016.7451781

17. Astafiev, S. V., Stanley, C. M., Shulman, G. L., & Corbetta, M. (2004). Extrastriate body area in human occipital cortex responds to the performance of motor actions. *Nature neuroscience, 7*(5), 542–548. https://doi.org/10.1038/nn1241

18. Aymerich-Franch, L., & Ferrer, I. (2022). Liaison, safeguard, and well-being: Analyzing the role of social robots during the COVID-19 pandemic. *Technology in Society, 70*, 101993.

19. Bach, P., Gunter, T. C., Knoblich, G., Prinz, W., & Friederici, A. D. (2009). N400-like negativities in action perception reflect the activation of two components of an action representation. *Social neuroscience, 4*(3), 212–232. https://doi.org/10.1080/17470910802362546

20. Baird, A., Parada-Cabaleiro, E., Hantke, S., Burkhardt, F., Cummins, N., & Schuller, B. (2018). The Perception and Analysis of the Likeability and Human Likeness of Synthesized Speech. *Interspeech*. 2863–2867. https://doi.org/10.21437/Interspeech.2018-1093

21. Baird, A. E., Jorgensen, S. H, Schuller, B., Cummins, N., Hantke, S., & Parada-Cabaliero, E. (2018). The Perception of Vocal Traits in Synthesized Voices: Age, Gender, and Human-Likeness. *Journal of the Audio Engineering Society*, *66*(4), 277-285. https://doi.org/10.17743/jaes.2018.0023

22. Balas, B., & Pacella, J. (2015). Artificial faces are harder to remember. Computers in Human Behavior, 52, 331–337. https://doi.org/10.1016/j.chb.2015.06.018

23. Barnhart, A. S., & Goldinger, S. D. (2013). Rotation reveals the importance of configural cues in handwritten word perception. *Psychonomic bulletin & review*, *20*(6), 1319–1326. https://doi.org/10.3758/s13423-013-0435-y

24. Barsics, C. (2014). Person recognition is easier from faces than from voices. *Psychologica Belgica, 54*(3), 244–254. https://doi.org/10.5334/pb.ap

25. Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology*, *25*(3), 281–316. https://doi.org/10.1006/cogp.1993.1007

26. Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2009). My Robotic Doppelganger - A Critical Look at the Uncanny Valley Theory. Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN2009, Toyama pp. 269-276.

27. Bastos, A. M., Lundqvist, M., Waite, A. S., Kopell, N., & Miller, E. K. (2020). Layer and rhythm specificity for predictive routing. *Proceedings of the National Academy of Sciences of the United States of America, 117*(49), 31459–31469. https://doi.org/10.1073/pnas.2014868117

28. Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

29. Baumann, O., & Belin, P. (2010). Perceptual scaling of voice identity: common dimensions for different vowels and speakers. *Psychological research, 74*(1), 110–120. https://doi.org/10.1007/s00426-008-0185-z

30. Bäuml K. H. (1994). Upright versus upside-down faces: how interface attractiveness varies with orientation. *Perception & psychophysics, 56*(2), 163–172. https://doi.org/10.3758/bf03213895

31. Becker-Asano, C.W., Ogawa, K., Nishio, S., & Ishiguro, H. (2010). Exploring the uncanny valley with Geminoid HI-1 in a real-world application. *IADIS Internaional Conference Interfaces and Human Computer Interaction*. 16302861.

32. Belin, P., Bestelmeyer, P. E., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British journal of psychology, 102*(4), 711–725. https://doi.org/10.1111/j.2044-8295.2011.02041.x

33. Benjamin, R., & Heine, S. J. (2023). From Freud to Android: Constructing a Scale of Uncanny Feelings. *Journal of personality assessment, 105*(1), 121–133. https://doi.org/10.1080/00223891.2022.2048842

34. Bilotta, E., Ariccio, S., Leone, L., & Bonaiuto, M. (2019). The Cumulative Risk Model to encompass perceived urban safety and well-being. https://doi.org/10.13135/2384-8677/3379

35. Björnström, L. E., Hills, C., Hanif, H., & Barton, J. J. (2014). Visual word expertise: a study of inversion and the word-length effect, with perceptual transforms. *Perception, 43*(5), 438–450. https://doi.org/10.1068/p7698

36. Bombari, D., Schmid, P. C., Schmid Mast, M., Birri, S., Mast, F. W., & Lobmaier, J. S. (2013). Emotion recognition: the role of featural and configural face information. *Quarterly journal of experimental psychology*, *66*(12), 2426–2442. https://doi.org/10.1080/17470218.2013.789065

37. Bond, C. F., Omar, A., Pitre, U., Lashley, B. R., Skaggs, L. M., & Kirk, C. T. (1992). Fishy-looking liars: Deception judgment from expectancy violation. *Journal of Personality and Social Psychology, 63*(6), 969–977. https://doi.org/10.1037/0022-3514.63.6.969

38. Boomsma, C., & Steg, L. (2014). Feeling Safe in the Dark. *Environment and Behavior*, *46*, 193– 212. https://doi.org/10.1177/0013916512453838

39. Bould, E., & Morris, N. (2008). Role of motion signals in recognizing subtle facial expressions of emotion. *British Journal of Psychology, 99*(2), 167–189. https://doi.org/10.1348/000712607X206702

40. Brandman, T., & Yovel, G. (2016). Bodies are Represented as Wholes Rather Than Their Sum of Parts in the Occipital-Temporal Cortex. *Cerebral cortex*, *26*(2), 530–543. https://doi.org/10.1093/cercor/bhu205

41. Broekens, J., Heerink, M., & Rosendal, H. (2009). Assistive social robots in elderly care: A review. *Gerontechnology, 8*(2), 94–103. https://doi.org/10.4017/gt.2009.08.02.002.00

42. Brosschot, J. F., Verkuil, B., & Thayer, J. F. (2018). Generalized Unsafety Theory of Stress: Unsafe Environments and Conditions, and the Default Stress Response. *International Journal of Environmental Research and Public Health, 15*(3), 464. https://doi.org/10.3390/ijerph15030464

43. Bukach, C. M., Gauthier, I., Tarr, M. J., Kadlec, H., Barth, S., Ryan, E., Turpin, J., & Bub, D. N. (2012). Does acquisition of Greeble expertise in prosopagnosia rule out a domain-general deficit?. *Neuropsychologia, 50*(2), 289–304. https://doi.org/10.1016/j.neuropsychologia.2011.11.023

44. 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, https://www.10.1016/j.chb.2012.11.02 1

45. Burleigh, T. J., & Schoenherr, J. R. (2015). A reappraisal of the uncanny valley: Categorical perception or frequency-based sensitization? *Frontiers in Psychology, 5*, Article 1488. https://doi.org/10.3389/fpsyg.2014.01488

46. Calder, A. J., Keane, J., Manes, F., Antoun, N., & Young, A. W. (2000). Impaired recognition and experience of disgust following brain injury. *Nature Neuroscience, 3*(11), 1077–1078. https://doi.org/10.1038/80586

47. Calder, A. J., & Jansen, J. (2005). Configural coding of facial expressions: The impact of inversion and photographic negative. *Visual Cognition, 12*(3), 495–518. https://doi.org/10.1080/13506280444000418

48. Calvo, M. G., Nummenmaa, L., & Avero, P. (2008). Visual search of emotional faces: Eye-movement assessment of component processes. *Experimental Psychology, 55*(6), 359–370. https://doi.org/10.1027/1618-3169.55.6.359

49. Cameron, L., Rutland, A., Brown, R., & Douch, R. (2006). Changing Children's Intergroup Attitudes Toward Refugees: Testing Different Models of Extended Contact. *Child Development, 77*(5), 1208–1219. https://doi.org/10.1111/j.1467-8624.2006.00929.x

50. Carbon, C.-C., & Leder, H. (2006). When faces are heads: View-dependent recognition of faces altered relationally or componentially. *Swiss Journal of Psychology / Schweizerische Zeitschrift für Psychologie / Revue Suisse de Psychologie, 65*(4), 245–252. https://doi.org/10.1024/1421-0185.65.4.245

51. Carbon, C. C., Schweinberger, S. R., Kaufmann, J. M., & Leder, H. (2005). The Thatcher illusion seen by the brain: an event-related brain potentials study. Brain research. *Cognitive brain research, 24*(3), 544–555. https://doi.org/10.1016/j.cogbrainres.2005.03.008

52. Carr, E. W., Hofree, G., Sheldon, K., Saygin, A. P., & Winkielman, P. (2017). Is that a human? Categorization (dis)fluency drives evaluations of agents ambiguous on human-likeness. Journal of experimental psychology. *Human Perception and Performance, 43*(4), 651–666. https://doi.org/10.1037/xhp0000304

53. Carroll, J. M., Yik, M. S. M., Russell, J. A., & Barrett, L. F. (1999). On the psychometric principles of affect. *Review of General Psychology, 3*(1), 14–22. https://doi.org/10.1037/1089-2680.3.1.14

54. Cassia, V. M., Kuefner, D., Westerlund, A., & Nelson, C. A. (2006). A behavioural and ERP investigation of 3-month-olds' face preferences. *Neuropsychologia, 44*(11), 2113–2125. https://doi.org/10.1016/j.neuropsychologia.2005.11.014

55. Chang, H.-H. (2013). Wayfinding strategies and tourist anxiety in unfamiliar destinations. *An International Journal of Tourism Space, Place, and Environment, 15*, 529–550. https://doi.org/10.1080/14616688.2012.726270

56. Chapman, H. A., & Anderson, A. K. (2012). Understanding disgust. *Annals of the New York Academy of Sciences, 1251*, 62–76. https://doi.org/10.1111/j.1749-6632.2011.06369.x

57. Chattopadhyay, D., & MacDorman, K. F. (2016). Familiar faces rendered strange: Why inconsistent realism drives characters into the uncanny valley. *Journal of Vision, 16*(11), 7. https://doi.org/10.1167%2F16.11.7

58. Cheetham, M., Suter, P., & Jäncke, L. (2011). The human likeness dimension of the "uncanny valley hypothesis": behavioral and functional MRI findings. *Frontiers in Human Neuroscience, 5*, 126. https://doi.org/10.3389/fnhum.2011.00126

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

60. Cheetham, M., Wu, L., Pauli, P., & Jancke, L. (2015). Arousal, valence, and the uncanny valley: psychophysiological and self-report findings. *Frontiers in Psychology, 6*, 981. https://doi.org/10.3389/fpsyg.2015.00981

61. Jóhannesson, Ó. I., Thornton, I. M., Smith, I. J., Chetverikov, A., & Kristjánsson, Á. (2016). Visual Foraging With Fingers and Eye Gaze. *i-Perception, 7*(2), 2041669516637279. https://doi.org/10.1177/2041669516637279

62. Cheung, O. S., & Gauthier, I. (2014). Visual appearance interacts with conceptual knowledge in object recognition. *Frontiers in Psychology, 5*, Article 793. https://doi.org/10.3389/fpsyg.2014.00793

63. Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology, 55*, 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015

64. Ciechanowski, L., Przegalinska, A.K., Magnuski, M., & Gloor, P.A. (2018). In the shades of the uncanny valley: An experimental study of human-chatbot interaction. *Future Generations Computer Systems, 92*, 539-548. https://doi.org/10.1016/j.future.2018.01.055

65. Coburn, A., Vartanian, O., Kenett, Y. N., Nadal, M., Hartung, F., Hayn-Leichsenring, G., Navarrete, G., González-Mora, J. L., & Chatterjee, A. (2020). Psychological and neural responses to architectural interiors. *Cortex; a journal devoted to the study of the nervous system and behavior, 126*, 217–241. https://doi.org/10.1016/j.cortex.2020.01.009

66. Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

67. Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

68. Colombatto, C., & McCarthy, G. (2017). The Effects of Face Inversion and Face Race on the P100 ERP. *Journal of cognitive neuroscience, 29*(4), 664–676. https://doi.org/10.1162/jocn_a_01079

69. Conde, K., & Pina, S.A. (2014). Urban Dimensions For Neighborhoods With Higher Environmental Value. *WIT Transactions on Ecology and the Environment, 191*, 415–426. https://doi.org/10.2495/SC140351

70. Connolly, H. L., Young, A. W., & Lewis, G. J. (2021). Face perception across the adult lifespan: evidence for age-related changes independent of general intelligence.

Cognition & emotion, 35(5), 890–901. https://doi.org/10.1080/02699931.2021.1901657

71. Conway, A., Brady, N., & Misra, K. (2017). Holistic word processing in dyslexia. *PloS one, 12*(11), e0187326. https://doi.org/10.1371/journal.pone.0187326

72. Costa, P. T., & McCrae, R. R. (1992). The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders, 6*(4), 343–359. https://doi.org/10.1521/pedi.1992.6.4.343

73. Crookes, K., Ewing, L., Gildenhuys, J. D., Kloth, N., Hayward, W. G., Oxner, M., Pond, S., & Rhodes, G. (2015). How Well Do Computer-Generated Faces Tap Face Expertise?. *PloS one, 10*(11), e0141353. https://doi.org/10.1371/journal.pone.0141353

74. Cunningham, A. B., & Schreibman, L. (2008). Stereotypy in Autism: The Importance of Function. *Research in autism spectrum disorders, 2*(3), 469–479. https://doi.org/10.1016/j.rasd.2007.09.006

75. Curtis, V., Aunger, R., & Rabie, T. (2004). Evidence that disgust evolved to protect from risk of disease. *Biological sciences, 271*, 131–S133. https://doi.org/10.1098/rsbl.2003.0144

76. Davies, N. (2016). Can robots handle your healthcare? *Engineering & Technology, 11*(9), 58–61. https://doi.org/10.1049/et.2016.0907

77. Davis, S. (1971). Acoustic characteristics of normal and pathological voices. *Speech and Language, 1*, 271–335. https://doi.org/10.1016/B978-0-12-608601-0.50010-3

78. Dawe, J., Sutherland, C., Barco, A., & Broadbent, E. (2019). Can social robots help children in healthcare contexts? A scoping review. BMJ paediatrics open, 3(1), e000371. https://doi.org/10.1136/bmjpo-2018-000371

79. de Gelder, B., & Van den Stock, J. (2011). The Bodily Expressive Action Stimulus Test (BEAST). Construction and Validation of a Stimulus Basis for Measuring Perception of Whole Body Expression of Emotions. *Frontiers in Pasychology*, *2*, 181. https://doi.org/10.3389/fpsyg.2011.00181

80. Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. Trends in cognitive sciences, 15(6), 254–262. https://doi.org/10.1016/j.tics.2011.04.003

81. Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. (2018). The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. PLoS biology, 16(3), e2004103. https://doi.org/10.1371/journal.pbio.2004103

82. Derntl, B., Seidel, E. M., Kainz, E., & Carbon, C. C. (2009). Recognition of emotional expressions is affected by inversion and presentation time. Perception, 38(12), 1849–1862. https://doi.org/10.1068/p6448

83. Destephe, M., Brandao, M., Kishi, T., Zecca, M., Hashimoto, K., & Takanishi, A. (2015). Walking in the uncanny valley: importance of the attractiveness on the acceptance of a robot as a working partner. Frontiers in psychology, 6, 204. https://doi.org/10.3389/fpsyg.2015.00204

84. Deska, J. C., Almaraz, S. M., & Hugenberg, K. (2017). Of mannequins and men: Ascriptions of mind in faces are bounded by perceptual and processing similarities to human faces. Social Psychological and Personality Science, 8(2), 183–190. https://doi.org/10.1177/1948550616671404

85. DeVries, A. C., Glasper, E. R., & Detillion, C. E. (2003). Social modulation of stress responses. Physiology & behavior, 79(3), 399–407. https://doi.org/10.1016/s0031-9384(03)00152-5

86. de Vries, M., Holland, R. W., Chenier, T., Starr, M. J., & Winkielman, P. (2010). Happiness cools the warm glow of familiarity: Psychophysiological evidence that mood modulates the familiarity-affect link. Psychological Science, 21(3), 321–328. https://doi.org/10.1177/0956797609359878

87. Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. Journal of Experimental Psychology: General, 115(2), 107–117. https://doi.org/10.1037/0096-3445.115.2.107

88. Diel, A., & Lewis, M. (2022). Structural deviations drive an uncanny valley of physical places. *Journal of Environmental Psychology*, *82*, 101844. https://doi.org/10.1016/j.jenvp.2022.101844

89. Diel, A., & Lewis, M. (2022). Familiarity, orientation, and realism increase face uncanniness by sensitizing to facial distortions. *Journal of Vision*, *22*(4), 14. https://doi.org/10.1167/jov.22.4.14

90. Diel, A., & Lewis, M. (2022). The deviation-from-familiarity effect: Expertise increases uncanniness of deviating exemplars. *PloS one*, *17*(9), e0273861. https://doi.org/10.1371/journal.pone.0273861

91. Diel, A., & Lewis, M. (2022). The uncanniness of written text is explained by configural deviation and not by processing disfluency. *Perception*, 3010066221114436. Advance online publication. https://doi.org/10.1177/03010066221114436

92. Diel, A., & MacDorman, K. F. (2021). Creepy cats and strange high houses: Support for configural processing in testing predictions of nine uncanny valley theories. Journal of Vision, 21(4), Article 1. https://doi.org/10.1167/jov.21.4.1

93. Diel, A., Weigelt, S., & Macdorman, K. F. (2022). A Meta-analysis of the Uncanny Valley's Independent and Dependent Variables. *ACM Transactions on Human-Robot Interaction*, *11*(1), 1-33. https://doi.org/10.1145/3470742

94. Dien J. (2009). The neurocognitive basis of reading single words as seen through early latency ERPs: a model of converging pathways. Biological psychology, 80(1), 10–22. https://doi.org/10.1016/j.biopsycho.2008.04.013

95. Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, *41*, 417–440. https://doi.org/10.1146/annurev.ps.41.020190.002221

96. Di Natale, A. F., Simonetti, M. E., La Rocca, S., & Bricolo, E. (2023). Uncanny valley effect: A qualitative synthesis of empirical research to assess the suitability of using virtual faces in psychological research. *Computers in Human Behavior*, *10*, 100288. https://doi.org/10.1016/j.chbr.2023.100288

97. Doherty-Sneddon, G., Whittle, L., & Riby, D. M. (2013). Gaze aversion during social style interactions in autism spectrum disorder and Williams syndrome. Research in developmental disabilities, 34(1), 616–626. https://doi.org/10.1016/j.ridd.2012.09.022

98. Donnelly, N., Zürcher, N. R., Cornes, K., Snyder, J., Naik, P., Hadwin, J., & Hadjikhani, N. (2011). Discriminating grotesque from typical faces: evidence from the Thatcher illusion. PloS one, 6(8), e23340. https://doi.org/10.1371/journal.pone.0023340

99. Dotsch, R., Hassin, R.R., & Todorov, A. (2016). Statistical learning shapes face evaluation. Nature Human Behaviour, 1. https://doi.org/10.1038/s41562-016-0001

100. Durand, K., Gallay, M., Seigneuric, A., Robichon, F., & Baudouin, J. Y. (2007). The development of facial emotion recognition: the role of configural

information. Journal of experimental child psychology, 97(1), 14–27. https://doi.org/10.1016/j.jecp.2006.12.001

101. Eadie, T. L., Rajabzadeh, R., Isetti, D. D., Nevdahl, M. T., & Baylor, C. R. (2017). The Effect of Information and Severity on Perception of Speakers With Adductor Spasmodic Dysphonia. American journal of speech-language pathology, 26(2), 327–341. https://doi.org/10.1044/2016_AJSLP-15-0191

102. Eddington, C. M., & Tokowicz, N. (2015). How meaning similarity influences ambiguous word processing: the current state of the literature. Psychonomic bulletin & review, 22(1), 13–37. https://doi.org/10.3758/s13423-014-0665-7

103. Eimer M. (2011). The face-sensitivity of the n170 component. Frontiers in human neuroscience, 5, 119. https://doi.org/10.3389/fnhum.2011.00119

104. Epstein, R. A., & Baker, C. I. (2019). Scene Perception in the Human Brain. Annual review of vision science, 5, 373–397. https://doi.org/10.1146/annurev-vision-091718-014809

105. Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. Nature, 392(6676), 598–601. https://doi.org/10.1038/33402

106. Fallshore, M., & Bartholow, J. (2003). Recognition of emotion from inverted schematic drawings of faces. Perceptual and motor skills, 96(1), 236–244. https://doi.org/10.2466/pms.2003.96.1.236

107. Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? Journal of Experimental Psychology: Human Perception and Performance, 21(3), 628–634. https://doi.org/10.1037/0096-1523.21.3.628

108. Fattal, C., Cossin, I., Pain, F., Haize, E., Marissael, C., Schmutz, S., & Ocnarescu, I. (2020). Perspectives on usability and accessibility of an autonomous

humanoid robot living with elderly people. *Disability and Rehabilitation: Assistive Technology*, *17*(4), 418–430. https://doi.org/10.1080/17483107.2020.1786732

109. Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior research methods, 39(2), 175–191. https://doi.org/10.3758/bf03193146

110. Faulkner, J., Schaller, M., Park, J. H., & Duncan, L. A. (2004). Evolved Disease-Avoidance Mechanisms and Contemporary Xenophobic Attitudes. Group Processes & Intergroup Relations, 7(4), 333–353. https://doi.org/10.1177/1368430204046142

111. Ferrari, F., Paladino, M.P. & Jetten, J. Blurring Human–Machine Distinctions: Anthropomorphic Appearance in Social Robots as a Threat to Human Distinctiveness. International Journal of Social Robotics 8, 287–302 (2016). https://doi.org/10.1007/s12369-016-0338-y

112. Ferrey, A. E., Hughes, N. D., Simkin, S., Locock, L., Stewart, A., Kapur, N., Gunnell, D., & Hawton, K. (2016). The impact of self-harm by young people on parents and families: a qualitative study. BMJ open, 6(1), e009631. https://doi.org/10.1136/bmjopen-2015-009631

113. Festinger, L. (1957). A theory of cognitive dissonance. Stanford University Press.

114. Fincher, K. M., & Tetlock, P. E. (2016). Perceptual dehumanization of faces is activated by norm violations and facilitates norm enforcement. Journal of Experimental Psychology: General, 145(2), 131–146. https://doi.org/10.1037/xge0000132

115. Fincher, K. M., Tetlock, P. E., & Morris, M. W. (2017). Interfacing with faces: Perceptual humanization and dehumanization. Current Directions in Psychological Science, 26(3), 288–293. https://doi.org/10.1177/0963721417705390

116. Fiser, A., Mahringer, D., Oyibo, H. K., Petersen, A. V., Leinweber, M., & Keller, G. B. (2016). Experience-dependent spatial expectations in mouse visual cortex. Nature neuroscience, 19(12), 1658–1664. https://doi.org/10.1038/nn.4385

117. Fisher, M. (2016). The Weird and the Eerie. Repeater.

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

119. Freud, S. (1919/2003). The uncanny [das unheimliche] (D. McLintock, Trans.). New York: Penguin.

120. Friston K. (2010). The free-energy principle: a unified brain theory?. Nature reviews. Neuroscience, 11(2), 127–138. https://doi.org/10.1038/nrn2787

121. Friston, K. J. & Kiebel, S. Predictive coding: A free-energy formulation. Predictions in the Brain 231–246 (2011). https://www.10.1093/acprof:oso/9780195395518.003.0076

122. Fujimura, T. & Umemura, H. (2018). Development and validation of a facial expression database based on the dimensional and categorical model of emotions. *Cognition & Emotion*, *32*(8), 1663–1670. https://doi.org/10.1080/02699931.2017.1419936

123. Gaudrain, E., Li, S., Ban, V.S., & Patterson, R.D. (2009). The role of glottal pulse rate and vocal tract length in the perception of speaker identity. Interspeech. https://doi.org/10.6084/M9.FIGSHARE.870509.V1

124. Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nature neuroscience, 2(6), 568–573. https://doi.org/10.1038/9224

125. Gauthier, I., & Nelson, C. A. (2001). The development of face expertise. Current opinion in neurobiology, 11(2), 219–224. https://doi.org/10.1016/s0959-4388(00)00200-2

126. Gauthier I., Wong A. C.-N., Hayward W. G., Cheung O. S. (2006). Font tuning associated with expertise in letter perception. Perception, 35(4), 541–559. https://doi.org/10.1068/p5313

127. Goffman, E. (1963). Stigma. Notes on the Management of Spoiled Identity. London: Penguin Books.

128. Goldberg, L. R. (1999). A Broad-Bandwidth, Public Domain Personality Inventory Measuring the Lower-Level Facets of Several Five-Factor Models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), Personality Psychology in Europe, Vol. 7 (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.

129. Gollwitzer, A., Marshall, J., & Bargh, J. A. (2020). Pattern deviancy aversion predicts prejudice via a dislike of statistical minorities. Journal of Experimental Psychology: General, 149(5), 828–854. https://doi.org/10.1037/xge0000682

130. Gollwitzer, A., Marshall, J., Wang, Y., & Bargh, J. A. (2017). Relating pattern deviancy aversion to stigma and prejudice. Nature human behaviour, 1(12), 920–927. https://doi.org/10.1038/s41562-017-0243-x

131. Gollwitzer, A., Martel, C., Heinecke, A., & Bargh, J. A. (2022). Deviancy Aversion and Social Norms. Personality & social psychology bulletin,

1461672221131378. Advance online publication. https://doi.org/10.1177/01461672221131378

132. Lawick-Goodall, J.V. (1968). The Behaviour of Free-living Chimpanzees in the Gombe Stream Reserve. Animal Behaviour Monographs, 1, 161-311. https://doi.org/10.1016/S0066-1856%2868%2980003-2

133. Goren, D., & Wilson, H. R. (2006). Quantifying facial expression recognition across viewing conditions. Vision Research, 46(8-9), 1253–1262. https://doi.org/10.1016/j.visres.2005.10.028

134. Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: mind perception and the uncanny valley. Cognition, 125(1), 125–130. https://doi.org/10.1016/j.cognition.2012.06.007

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

136. Greenberg, J., Pyszczynski, T., Solomon, S. (1986). The Causes and Consequences of a Need for Self-Esteem: A Terror Management Theory. In: Baumeister, R.F. (eds) Public Self and Private Self. Springer Series in Social Psychology. Springer, New York, NY. https://doi.org/10.1007/978-1-4613-9564-5_10

137. Grillon, C., Pellowski, M., Merikangas, K. R., & Davis, M. (1997). Darkness facilitates the acoustic startle reflex in humans. Biological psychiatry, 42(6), 453–460. https://doi.org/10.1016/S0006-3223(96)00466-0

138. Güçlütürk, Y., Jacobs, R. H., & van Lier, R. (2016). Liking versus Complexity: Decomposing the Inverted U-curve. Frontiers in human neuroscience, 10, 112. https://doi.org/10.3389/fnhum.2016.00112

139. Hahn, A. C., Jantzen, K. J., & Symons, L. A. (2012). Thatcherization impacts the processing of own-race faces more so than other-race faces: an ERP study. Social neuroscience, 7(2), 113–125. https://doi.org/10.1080/17470919.2011.583080

140. Haidt, J., McCauley, C., & Rozin, P. (1994). Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors. Personality and Individual Differences, 16(5), 701–713. https://doi.org/10.1016/0191-8869(94)90212-7

141. Halberstadt, J., & Winkielman, P. (2013). When good blends go bad: How fluency can explain when we like and dislike ambiguity. In C. Unkelbach & R. Greifender (Eds.), The experience of thinking: How the fluency of mental processes influences cognition and behaviour (pp. 133–150). Psychology Press.

142. Halberstadt, J., & Winkielman, P. (2014). Easy on the eyes, or hard to categorize: Classification difficulty decreases the appeal of facial blends. Journal of Experimental Social Psychology, 50, 175–183. https://doi.org/10.1016/j.jesp.2013.08.004

143. Halit, H., de Haan, M., & Johnson, M. H. (2003). Cortical specialisation for face processing: face-sensitive event-related potential components in 3- and 12-month-old infants. NeuroImage, 19(3), 1180–1193. https://doi.org/10.1016/s1053-8119(03)00076-4

144. Hanson, D. (2006). Exploring the Aesthetic Range for Humanoid Robots.

145. Haq, S., & Zimring, C. (2003). Just down the road a piece: The development of topological knowledge of building layouts. Environment and Behavior, 35(1), 132–160. https://doi.org/10.1177/0013916502238868

146. Hartung, F., Jamrozik, A., Rosen, M. E., Aguirre, G., Sarwer, D. B., & Chatterjee, A. (2019). Behavioural and Neural Responses to Facial Disfigurement. Scientific reports, 9(1), 8021. https://doi.org/10.1038/s41598-019-44408-8

147. Herrmann, M. J., Ehlis, A. C., Ellgring, H., & Fallgatter, A. J. (2005). Early stages (P100) of face perception in humans as measured with event-related potentials (ERPs). Journal of neural transmission (Vienna, Austria : 1996), 112(8), 1073–1081. https://doi.org/10.1007/s00702-004-0250-8

148. Hillis, A. E., Newhart, M., Heidler, J., Barker, P., Herskovits, E., & Degaonkar, M. (2005). The roles of the "visual word form area" in reading. NeuroImage, 24(2), 548–559. https://doi.org/10.1016/j.neuroimage.2004.08.026

149. Hino, Y., Lupker, S. J., & Pexman, P. M. (2002). Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: Interactions between orthography, phonology, and semantics. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28(4), 686–713. https://doi.org/10.1037/0278-7393.28.4.686

150. Hölscher, C., Meilinger, T., Vrachliotis, G., Brösamle, M., & Knauff, M. (2006). Up the down staircase: Wayfinding strategies in multi-level buildings. Journal of Environmental Psychology, 26(4), 284–299. https://doi.org/10.1016/j.jenvp.2006.09.002

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

152. Ho, C.-C., & MacDorman, K. F. (2017). Measuring the uncanny valley effect: Refinements to indices for perceived humanness, attractiveness, and eeriness.

International Journal of Social Robotics, 9(1), 129–139. https://doi.org/10.1007/s12369-016-0380-9

153. Ho, C., Macdorman, K.F., & Pramono, Z.A. (2008). Human emotion and the uncanny valley: A GLM, MDS, and Isomap analysis of robot video ratings. 2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI), 169-176.

154. Huang, W-J., Xiao, H., & Wang, S. (2018). Airports as liminal space. *Annals of Tourism Research, 70*, 1–13. https://doi.org/10.1016/j.annals.2018.02.003

155. Hugenberg, K., Young, S., Rydell, R. J., Almaraz, S., Stanko, K. A., See, P. E., & Wilson, J. P. (2016). The face of humanity: Configural face processing influences ascriptions of humanness. Social Psychological and Personality Science, 7(2), 167–175. https://doi.org/10.1177/1948550615609734

156. Iftikhar, H., Shah, P.B., & Luximon, Y. (2020). Human wayfinding behaviour and metrics in complex environments: a systematic literature review. Architectural Science Review, 64, 452 - 463. https://doi.org/10.1080/00038628.2020.1777386

157. Imamoglu, Ç. (2000). Complexity, liking and familiarity: Architecture and non-architecture Turkish students' assessments of traditional and modern house facades. Journal of Environmental Psychology, 20(1), 5–16. https://doi.org/10.1006/jevp.1999.0155

158. Itier, R. J., & Taylor, M. J. (2002). Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: a repetition study using ERPs. NeuroImage, 15(2), 353–372. https://doi.org/10.1006/nimg.2001.0982

159. Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. Visual Cognition, 12(6), 1093–1123. https://doi.org/10.1080/13506280444000661

160. Jakesch, M., Leder, H., & Forster, M. (2013). Image ambiguity and fluency. PloS one, 8(9), e74084. https://doi.org/10.1371/journal.pone.0074084

161. Jamrozik, A., Oraa Ali, M., Sarwer, D. B., & Chatterjee, A. (2019). More than skin deep: Judgments of individuals with facial disfigurement. Psychology of Aesthetics, Creativity, and the Arts, 13(1), 117–129. https://doi.org/10.1037/aca0000147

162. Jaśkiewicz, M., & Besta, T. (2014). Is Easy Access Related to Better Life? Walkability and Overlapping of Personal and Communal Identity as Predictors of Quality of Life. Applied research in quality of life, 9(3), 505–516. https://doi.org/10.1007/s11482-013-9246-6

163. Jemel, B., George, N., Olivares, E., Fiori, N., & Renault, B. (1999). Event-related potentials to structural familiar face incongruity processing. Psychophysiology, 36(4), 437–452. https://doi.org/10.1017/S0048577299970853

164. Jentsch, E. (1906/1997). On the psychology of the uncanny (R. Sellars, Trans.). Angelaki, 2(1), 7–16. http://dx.doi.org/10.1080/09697259708571910.

165. Johnston, A., Brown, B. B., & Elson, R. (2021). Synchronous facial action binds dynamic facial features. Scientific reports, 11(1), 7191. https://doi.org/10.1038/s41598-021-86725-x

166. Jones, E. E., Farina, A., Hastorf, A. H., Markus, H., Miller, D. T., & Scott, R. A. (1984). Social stigma: The psychology of marked relationships. New York: Freeman.

167. Jung, N., Lee, M., & Choi, H. (2022). The uncanny valley effect for celebrity faces and celebrity-based avatars. *Science of Emotion and Sensibility*, *25*(1), 91–102. https://doi.org/10.14695/KJSOS.2022.25.1.91

168. Kanwisher, N. (2000). Domain specificity in face perception. Nature Neuroscience, 3(8), 759–763. https://doi.org/10.1038/77664

169. Kanwisher, N., & Moscovitch, M. (2000). The cognitive neuroscience of face processing: an introduction. Cognitive neuropsychology, 17(1), 1–11. https://doi.org/10.1080/026432900380454

170. Kaplan, S. (1987). Aesthetics, affect, and cognition: Environmental preference from an evolutionary perspective. Environment and Behavior, 19(1), 3–32. https://doi.org/10.1177/0013916587191001

171. Kaplan, F. (2004). Who is Afraid of the Humanoid? Investigating Cultural Differences in the Acceptance of Robots. International Journal of Humanoid Robotics, 1, 465-480. https://doi.org/10.1142/S0219843604000289

172. Kaplan, R., & Kaplan, S. (2011). Well-being, reasonableness, and the natural environment. Applied Psychology: Health and Well-Being, 3(3), 304–321. https://doi.org/10.1111/j.1758-0854.2011.01055.x

173. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and Improving the Image Quality of StyleGAN. *2020 IEEE/CFV Conference on Computer Vision and Pattern Recognition (CVPR)*, 8107–8116.

174. Kätsyri, J. (2018). Those virtual people all look the same to me: Computer-rendered faces elicit a higher false alarm rate than real human faces in a recognition memory task. Frontiers in Psychology, 9, Article 1362. https://doi.org/10.3389/fpsyg.2018.01362

175. 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. https://doi.org/10.3389/fpsyg.2015.00390

176. Kätsyri, J., de Gelder, B., & Takala, T. (2019). Virtual Faces Evoke Only a Weak Uncanny Valley Effect: An Empirical Investigation With Controlled Virtual Face Images. Perception, 48(10), 968–991. https://doi.org/10.1177/0301006619869134

177. Kawabe, T., Sasaki, K., Ihaya, K., & Yamada, Y. (2017). When categorization-based stranger avoidance explains the uncanny valley: A comment on MacDorman and Chattopadhyay (2016). Cognition, 161, 129–131. https://doi.org/10.1016/j.cognition.2016.09.001

178. Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, 3933-3936. https://doi.org/10.1109/ICASSP.2008.4518514

179. Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive Processing: A Canonical Cortical Computation. Neuron, 100(2), 424–435. https://doi.org/10.1016/j.neuron.2018.10.003

180. Keye, Z., Mingming, Z,, Tiantian, L., Wenbo, L., & Weiqi, H. (2017). The inversion effect of body recognition. *Advances in Psychological Science*, *27*(1), 27–36. https://doi.org/10.3724/SP.J.1042.2019.00027

181. Kim, D.-G. Kim, H.-Y., Kim, G., Jang, P.-S., Jung, W.-H., & Hyun, J.-S. (2016). Exploratory understanding of the uncanny valley phenomenon based on event-related potential measurement. Korean Society for Emotion and Sensibility, 19, 95–110. http://dx.doi.org/10.14695/KJSOS.2016.19.1.95

182. Kim, B., de Visser, E., & Phillips, E. (2022). Two uncanny valleys: Re-evaluating the uncanny valley across the full spectrum of real-world human-like

robots. *Computers in Human Behavior, 135,* 107340. https://doi.org/10.1016/j.chb.2022.107340

183. Kimura, M., & Yotsumoto, Y. (2018). Auditory traits of "own voice". PloS one, 13(6), e0199443. https://doi.org/10.1371/journal.pone.0199443

184. Kitchin, R. (1998). "Out of place," "knowing one's place": Space, power and the exclusion of disabled people. Disability & Society, 13(3), 343–356. https://doi.org/10.1080/09687599826678

185. Klauer, K. C., & Musch, J. (2003). Affective priming: Findings and theories. In J. Musch & K. C. Klauer (Eds.), The psychology of evaluation: Affective processes in cognition and emotion (pp. 7–49). Lawrence Erlbaum Associates Publishers.

186. Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. Journal of Neurolinguistics, 20(1), 1–24. https://doi.org/10.1016/j.jneuroling.2006.02.001

187. Koch, J. A., Bolderdijk, J. W., & van Ittersum, K. (2021). Disgusting? No, just deviating from internalized norms. Understanding consumer skepticism toward sustainable food alternatives. *Journal of Enviornmental Psychology, 76,* 101645. https://doi.org/10.1016/j.jenvp.2021.101645

188. Koschate, M., Potter, R., Bremner, P.A., & Levine, M. (2016). Overcoming the uncanny valley: Displays of emotions reduce the uncanniness of humanlike robots. 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 359-366. https://doi.org/10.1109/HRI.2016.7451773

189. Kreiman, J. E., Auszmann, A. & Gerratt, B. R. (2018). What does it mean for a voice to be “normal?” The Journal of the Acoustical Society of America 143, 1820–1820.

190. Kreiman, J., & Gerratt, B. R. (2005). Perception of aperiodicity in pathological voice. *The Journal of the Acoustical Society of America, 117*, 2201–2211. https://doi.org/10.1121/1.1858351

191. Kreiman, J., Gerratt, B. R., Precoda, K., & Berke, G. S. (1992). Individual differences in voice quality perception. Journal of speech and hearing research, 35(3), 512–520. https://doi.org/10.1044/jshr.3503.512

192. Krumhuber, E. G., Skora, L., Küster, D., & Fou, L. (2017). A review of dynamic datasets for facial expression research. Emotion Review, 9(3), 280–292. https://doi.org/10.1177/1754073916670022

193. Krumhuber, E. G., Tamarit, L., Roesch, E. B., & Scherer, K. R. (2012). FACSGen 2.0 animation software: Generating three-dimensional FACS-valid facial expressions for emotion research. Emotion, 12(2), 351–363. https://doi.org/10.1037/a0026632

194. Krypotos, A. M., Blanken, T. F., Arnaudova, I., Matzke, D., & Beckers, T. (2017). A Primer on Bayesian Analysis for Experimental Psychopathologists. Journal of experimental psychopathology, 8(2), 140–157. https://doi.org/10.5127/jep.057316

195. Kühne, K., Fischer, M. H., & Zhou, Y. (2020). The Human Takes It All: Humanlike Synthesized Voices Are Perceived as Less Eerie and More Likable. Evidence From a Subjective Ratings Study. Frontiers in neurorobotics, 14, 593732. https://doi.org/10.3389/fnbot.2020.593732

196. Kumazaki, H., Warren, Z., Muramatsu, T., Yoshikawa, Y., Matsumoto, Y., Miyao, M., Nakano, M., Mizushima, S., Wakita, Y., Ishiguro, H., Mimura, M., Minabe, Y., & Kikuchi, M. (2017). A pilot study for robot appearance preferences among high-functioning individuals with autism spectrum disorder: Implications for

therapeutic use. *PLOS ONE*, *12*(10), e0186581. https://doi.org/10.1371/journal.pone.0186581

197. Kurzban, R., & Leary, M. R. (2001). Evolutionary origins of stigmatization: The functions of social exclusion. Psychological Bulletin, 127(2), 187–208. https://doi.org/10.1037/0033-2909.127.2.187

198. Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. Science, 207(4427), 203–205. https://doi.org/10.1126/science.7350657

199. Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software, 82(13), 1–26. https://doi.org/10.18637/jss.v082.i13

200. Lamont, A. C., Stewart-Williams, S., & Podd, J. (2005). Face recognition and aging: effects of target age and memory load. Memory & cognition, 33(6), 1017–1024. https://doi.org/10.3758/bf03193209

201. Langer, M., & König, C. J. (2018). Introducing and testing the Creepiness of Situation Scale (CRoSS). Frontiers in Psychology, 9, Article 2220. https://doi.org/10.3389/fpsyg.2018.02220

202. Langer, M., König, C. J., & Fitili, A. (2018). Information as a double-edged sword: The role of computer experience and information on applicant reactions towards novel technologies for personnel selection. Computers in Human Behavior, 81, 19–30. https://doi.org/10.1016/j.chb.2017.11.036

203. Langer, M., König, C. J., & Papathanasiou, M. (2019). Highly automated job interviews: Acceptance under the influence of stakes. International Journal of Selection and Assessment, 27(3), 217–234. https://doi.org/10.1111/ijsa.12246

204. Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. Psychological Science, 1(2), 115–121. https://doi.org/10.1111/j.1467-9280.1990.tb00079.x

205. Latinus, M., VanRullen, R., & Taylor, M. J. (2010). Top-down and bottom-up modulation in processing bimodal face/voice stimuli. BMC Neuroscience, 11, Article 36. https://doi.org/10.1186/1471-2202-11-36

206. Latinus, M., McAleer, P., Bestelmeyer, P. E., & Belin, P. (2013). Norm-based coding of voice identity in human auditory cortex. Current biology : CB, 23(12), 1075–1080. https://doi.org/10.1016/j.cub.2013.04.055

207. Laurence, P. G., Pinto, T. M., Rosa, A. T. F., & Macedo, E. C. (2018). Can a lexical decision task predict efficiency in the judgment of ambiguous sentences?. Psicologia, reflexao e critica : revista semestral do Departamento de Psicologia da UFRGS, 31(1), 13. https://doi.org/10.1186/s41155-018-0093-0

208. Law, G. U., Sinclair, S., & Fraser, N. (2007). Children's attitudes and behavioural intentions towards a peer with symptoms of ADHD: does the addition of a diagnostic label make a difference?. Journal of child health care : for professionals working with children in the hospital and community, 11(2), 98–111. https://doi.org/10.1177/1367493507076061

209. Leander, N. P., Chartrand, T. L., & Bargh, J. A. (2012). You give me the chills: Embodied reactions to inappropriate amounts of behavioral mimicry. Psychological Science, 23(7), 772–779. https://doi.org/10.1177/0956797611434535Leder, Goller, Forster, Schlageter, & Paul, 2017

210. Leder H. (1996). Line drawings of faces reduce configural processing. Perception, 25(3), 355–366. https://doi.org/10.1068/p250355

211. Lee, Kang and others, ' Development of Face Processing Expertise', in Andrew J. Calder and others (eds), Oxford Handbook of Face Perception, Oxford Library of Psychology (2011; online edn, Oxford Academic, 21 Nov. 2012), https://doi.org/10.1093/oxfordhb/9780199559053.013.0039, accessed 20 May 2023.

212. Levine, T. R., Anders, L. N., Banas, J., Baum, K. L., Endo, K., Hu, A. D. S., & Wong, N. C. H. (2000). Norms, expectations, and deception: A norm violation model of veracity judgments. Communication Monographs, 67(2), 123–137. https://doi.org/10.1080/03637750009376500

213. Lewis M. B. (2001). The lady's not for turning: rotation of the Thatcher illusion. Perception, 30(6), 769–774. https://doi.org/10.1068/p3174

214. Li, C. (2006). User preferences, information transactions and location-based services: A study of urban pedestrian wayfinding. *Computers, Environment, and Urban Systems, 30*(6), 726–740. https://doi.org/10.1016/j.compenvurbsys.2006.02.008

215. Lim, A., Young, R. L., & Brewer, N. (2022). Autistic adults may be erroneously perceived as deceptive and lacking credibility. Journal of Autism and Developmental Disorders, 52(2), 490–507. https://doi.org/10.1007/s10803-021-04963-4

216. Lischetzke, T., Izydorczyk, D., Hüller, C., & Appel, M. (2017). The topography of the uncanny valley and individuals' need for structure: A nonlinear mixed effects analysis. *Journal of Research in Personality, 68,* 96–113. https://doi.org/10.1016/j.jrp.2017.02.001

217. Logačev, P., & Vasishth, S. (2016). A Multiple-Channel Model of Task-Dependent Ambiguity Resolution in Sentence Comprehension. Cognitive science, 40(2), 266–298. https://doi.org/10.1111/cogs.12228

218. Löffler, D., Dörrenbächer, J., & Hassenzahl, M. (2020). The Uncanny Valley Effect in Zoomorphic Robots: The U-Shaped Relation Between Animal Likeness and Likeability. 2020 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 261-270. https://doi.org/10.1145/3319502.3374788

219. Löffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. Nature Neuroscience, 8(10), 1386–1390. https://doi.org/10.1038/nn1538

220. Looser, C. E., & Wheatley, T. (2010). The tipping point of animacy. How, when, and where we perceive life in a face. Psychological science, 21(12), 1854–1862. https://doi.org/10.1177/0956797610388044

221. Lu, V. N., Wirtz, J., Kunz, W. H., Paluch, S., Gruber, T., Martins, A., & Patterson, P. G. (2020). Service robots, customers and service employees: What can we learn from the academic literature and where are the gaps? Journal of Service Theory and Practice, 30(3), 361–391. https://doi.org/10.1108/JSTP-04-2019-0088

222. Luke S. G. (2017). Evaluating significance in linear mixed-effects models in R. Behavior research methods, 49(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

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

224. Ma, T., Sharifi, H., & Chattopadhyay, D. (2019). Virtual Humans in Health-Related Interventions: A Meta-Analysis. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3290607.3312853

225. MacDonald, A. P. (1970). Revised scale for ambiguity tolerance: Reliability and validity. Psychological Reports, 26(3), 791–798. https://doi.org/10.2466/pr0.1970.26.3.791

226. MacDonnell, R., Breidt, M., & Bülthoff, H. H. (2012). Render me real?: Investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics*, *31*(4), 1–11. https://doi.org/10.1145/2185520.2185587

227. MacDorman, K.F. (2005). Mortality salience and the uncanny valley. 5th IEEE-RAS International Conference on Humanoid Robots, 2005., 399-405. https://doi.org/10.1109/ICHR.2005.1573600

228. 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

229. MacDorman, K. F., & Entezari, S. O. (2015). Individual differences predict sensitivity to the uncanny valley. Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems, 16(2), 141–172. https://doi.org/10.1075/is.16.2.01mac

230. MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems, 7(3), 297–337. https://doi.org/10.1075/is.7.3.03mac

231. Mäkäräinen, M., Kätsyri, J., & Takala, T. (2014). Exaggerating facial expressions: A way to intensify emotion or a way to the uncanny valley? *Cognitive Computation*, *6*, 708–721. https://doi.org/10.1007/s12559-014-9273-0

232. Makino, H., & Komiyama, T. (2015). Learning enhances the relative impact of top-down processing in the visual cortex. Nature neuroscience, 18(8), 1116–1122. https://doi.org/10.1038/nn.4061

233. Manago, B., & Mize, T. D. (2022). The status and stigma consequences of mental illness labels, deviant behavior, and fear. Social science research, 105, 102690. https://doi.org/10.1016/j.ssresearch.2021.102690

234. Mangan, B. (2015). The uncanny valley as fringe experience. Interaction Studies, 16, 193-199. https://doi.org/10.1075/IS.16.2.05MAN

235. Mara, M., Appel, M., & Gnambs, T. (2022). Human-like robots and the uncanny valley: A meta-analysis of user responses based on the godspeed scales. *Zeitschrift für Psychologie*, *230*(1), 33–46. https://doi.org/10.1027/2151-2604/a000486

236. Martelli, M., Majaj, N. J., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. Journal of vision, 5(1), 58–70. https://doi.org/10.1167/5.1.6

237. Martinez A. M. (2017). Visual perception of facial expressions of emotion. Current opinion in psychology, 17, 27–33. https://doi.org/10.1016/j.copsyc.2017.06.009

238. Masschelein, A. (2012). The Unconcept. The Freudian Uncanny in late-twentieth-century theory. *State University of New York Press.*

239. 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

240. Mathur, M. B., Reichling, D. B., Lunardini, F., Geminiani, A., Antonietti, A., Ruijten, P. A. M., Levitan, C. A., Nave, G., Manfredi, D., Bessette-Symons, B., Szuts,

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. https://doi.org/10.1016/j.chb.2019.08.029

241. Matsuda, Y. T., Okamoto, Y., Ida, M., Okanoya, K., & Myowa-Yamakoshi, M. (2012). Infants prefer the faces of strangers or mothers to morphed faces: an uncanny valley between social novelty and familiarity. Biology letters, 8(5), 725–728. https://doi.org/10.1098/rsbl.2012.0346

242. Mattavelli, G., Sormaz, M., Flack, T., Asghar, A. U., Fan, S., Frey, J., Manssuer, L., Usten, D., Young, A. W., & Andrews, T. J. (2014). Neural responses to facial expressions support the role of the amygdala in processing threat. Social cognitive and affective neuroscience, 9(11), 1684–1689. https://doi.org/10.1093/scan/nst162

243. Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. Trends in Cognitive Sciences, 6(6), 255–260. https://doi.org/10.1016/S1364-6613(02)01903-4

244. Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: a comparison of language and faces. Developmental psychobiology, 56(2), 154–178. https://doi.org/10.1002/dev.21177

245. McGregor, I., Haji, R., & Kang, S.-J. (2008). Can ingroup affirmation relieve outgroup derogation? *Journal of Experimental Social Psychology*, *44*(5), 1395–1401. https://doi.org/10.1016/j.jesp.2008.06.001

246. Meiri, N., Schnapp, Z., Ankri, A., Nahmias, I., Raviv, A., Sagi, O., Hamad Saied, M., Konopnicki, M., & Pillar, G. (2017). Fear of clowns in hospitalized children: prospective experience. *European Journal of Pediatrics*, *176*(2), 269–272. https://doi.org/10.1007/s00431-016-2826-3

247. Michalareas, G., Vezoli, J., van Pelt, S., Schoffelen, J. M., Kennedy, H., & Fries, P. (2016). Alpha-Beta and Gamma Rhythms Subserve Feedback and Feedforward Influences among Human Visual Cortical Areas. Neuron, 89(2), 384–397. https://doi.org/10.1016/j.neuron.2015.12.018

248. Milivojevic, B., Clapp, W. C., Johnson, B. W., & Corballis, M. C. (2003). Turn that frown upside down: ERP effects of thatcherization of misorientated faces. Psychophysiology, 40(6), 967–978. https://doi.org/10.1111/1469-8986.00115

249. Miller, E.J., Foo, Y.Z., Mewton, P., & Dawel, A. (2023). How do people respond to computer-generated versus human faces? A systematic review and meta-analyses. *Computers in Human Behavior Reports*, *10*, 100283. https://doi.org/10.1016/j.chbr.2023.100283

250. McAndrew, F.T. (2020). The Psychology, Geography, and Architecture of Horror: How Places Creep Us Out. Evolutionary Studies in Imaginative Culture, 4, 47 - 62. https://doi.org/10.26613/esic.4.2.189

251. McAndrew, F. T., & Koehnke, S. S. (2016). On the nature of creepiness. New Ideas in Psychology, 43, 10–15. https://doi.org/10.1016/j.newideapsych.2016.03.003 McGlone & Tofighbakhsh, 2000

252. McKelvie S. J. (1995). Emotional expression in upside-down faces: evidence for configurational and componential processing. The British journal of social psychology, 34 (Pt 3), 325–334. https://doi.org/10.1111/j.2044-8309.1995.tb01067.x

253. McLean, R. A., Sanders, W. L., & Stroup, W. W. (1991). A Unified Approach to Mixed Linear Models. The American Statistician, 45(1), 54–64. https://doi.org/10.2307/2685241

254. Meah, L. F. S., & Moore, R. K. (2014). The uncanny valley: A focus on misaligned cues. In Beetz, M., Johnston, B., & Williams, M.-A. (Eds.), Social

Robotics: 6th International Conference (pp. 256–265). Lecture Notes in Computer Science, vol. 8755. Cham, Switzerland: Springer.

255. Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. Proceedings of the National Academy of Sciences of the United States of America, 108(48), 19401–19406. https://doi.org/10.1073/pnas.1112895108

256. Mitchell, W. J., Szerszen, 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. https://doi.org/10.1068/i0415

257. Moore, R. (2012). A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena. *Scientific Reports*, *2*, 864. https://doi.org/10.1038/srep00864

258. Moosa, M.M., & Ud-Dean, S.M. (2010). Danger Avoidance: An Evolutionary Explanation of Uncanny Valley. *Biological Theory, 5*, 12-14. https://doi.org/10.1162/BIOT_A_00016

259. Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. IEEE Robotics & Automation Magazine, 19, 98–100. https://doi.org/10.1109/MRA.2012.2192811

260. Morris, J. (2001). Impairment and Disability: Constructing an Ethics of Care That Promotes Human Rights. Hypatia, 16, 1 - 16. https://doi.org/10.1111/j.1527-2001.2001.tb00750.x

261. Mühlberger, A., Neumann, R., Wieser, M. J., & Pauli, P. (2008). The impact of changes in spatial distance on emotional responses. Emotion, 8(2), 192–198. https://doi.org/10.1037/1528-3542.8.2.192

262. Müller, B. C. N., Gao, X., Nijssen, S. R. R., & Damen, T. G. E. (2021). I, Robot: How Human Appearance and Mind Attribution Relate to the Perceived Danger of Robots. International Journal of Social Robotics, 13, 691–701. https://doi.org/10.1007/s12369-020-00663-8

263. Mustafa, M., Guthe, S., Tauscher, J.-P., Goesele, M., & Magnor, M. (2017). *How human am I? Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*.

264. Nakanishi, J., Kuramoto, I., Baba, J., Ogawa, K., Yoshikawa, Y., & Ishiguro, H. (2020). Continuous hospitality with social robots at a hotel. *SN Applied Sciences, 2*, 452. https://doi.org/10.1007/s42452-020-2192-7

265. Neuberg, S. L., & Cottrell, C. A. (2008). Managing the threats and opportunities afforded by human sociality. Group Dynamics: Theory, Research, and Practice, 12(1), 63–72. https://doi.org/10.1037/1089-2699.12.1.63

266. Neuberg, S. L., & Newsom, J. T. (1993). Personal need for structure: Individual differences in the desire for simpler structure. *Journal of Personality and Social Psychology*, *65*(1), 113–131. https://doi.org/10.1037/0022-3514.65.1.113

267. Neuhofer, B., Eger, R., Yu, J., & Celuch, K. (2021). Designing experiences in the age of human transformation: An analysis of Burning Man. *Annals of Tourism Research, 91,* 103310. https://doi.org/10.1016/j.annals.2021.103310

268. Nixon, E. (2001). The social competence of children with Attention Deficit Hyperactivity Disorder: A review of the literature. Child and Adolescent Mental Health, 6(4), 172–180. https://doi.org/10.1111/1475-3588.00342

269. Olaronke, I., Rhoda, I., & Janet, O. (2017). A framework for avoiding uncanny valley in healthcare. *IJBHTM*, *7*(1), 1–10.

270. Olatunji, B. O., Williams, N. L., Tolin, D. F., Abramowitz, J. S., Sawchuk, C. N., Lohr, J. M., & Elwood, L. S. (2007). The Disgust Scale: item analysis, factor structure, and suggestions for refinement. Psychological assessment, 19(3), 281–297. https://doi.org/10.1037/1040-3590.19.3.281

271. Olivares, E. I., & Iglesias, J. (2010). Brain potential correlates of the "internal features advantage" in face recognition. Biological psychology, 83(2), 133–142. https://doi.org/10.1016/j.biopsycho.2009.11.011

272. Olivares, E. I., Lage-Castellanos, A., Bobes, M. A., & Iglesias, J. (2018). Source reconstruction of brain potentials using Bayesian model averaging to analyze face intra-domain vs. face-occupation cross-domain processing. Frontiers in Integrative Neuroscience, 12, Article 12. https://doi.org/10.3389/fnint.2018.00012

273. Olivares, E. I., Iglesias, J., Saavedra, C., Trujillo-Barreto, N. J., & Valdés-Sosa, M. (2015). Brain Signals of Face Processing as Revealed by Event-Related Potentials. Behavioural neurology, 2015, 514361. https://doi.org/10.1155/2015/514361

274. Olivera-La Rosa, A., Villacampa, J., Corradi, G., & Ingram, G. P. D. (2021). The creepy, the bad and the ugly: Exploring perceptions of moral character and social desirability in uncanny faces. Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues. Advance online publication. https://doi.org/10.1007/s12144-021-01452-w

275. Oppenheimer, D. M. (2008). The secret life of fluency. Trends in Cognitive Sciences, 12(6), 237–241. https://doi.org/10.1016/j.tics.2008.02.014

276. Owen, H. E., Halberstadt, J., Carr, E. W., & Winkielman, P. (2016). Johnny Depp, Reconsidered: How Category-Relative Processing Fluency Determines the

Appeal of Gender Ambiguity. PloS one, 11(2), e0146328. https://doi.org/10.1371/journal.pone.0146328

277. Packer, C. and Pusey, A.E. (1984) Infanticide in Carnivores. In: Hausfater, G. and Hrdy, S.B., Eds., Infanticide: Comparative and Evolutionary Perspectives, Aldine, New York, 31-42.

278. Park, J. H., Faulkner, J., & Schaller, M. (2003). Evolved Disease-Avoidance Processes and Contemporary Anti-Social Behavior: Prejudicial Attitudes and Avoidance of People with Physical Disabilities. Journal of Nonverbal Behavior, 27(2), 65–87. https://doi.org/10.1023/A:1023910408854

279. Pascalis, O., de Martin de Viviés, X., Anzures, G., Quinn, P. C., Slater, A. M., Tanaka, J. W., & Lee, K. (2011). Development of face processing. Wiley interdisciplinary reviews. Cognitive science, 2(6), 666–675. https://doi.org/10.1002/wcs.146

280. Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. Journal of neurophysiology, 93(1), 603–608. https://doi.org/10.1152/jn.00513.2004

281. Pelli, D. G., Farell, B., & Moore, D. C. (2003). The remarkable inefficiency of word recognition. Nature, 423(6941), 752–756. https://doi.org/10.1038/nature01516

282. Perugini, M., Gallucci, M., & Costantini, G. (2014). Safeguard power as a protection against imprecise power estimates. Perspectives on Psychological Science, 9(3), 319–332. https://doi.org/10.1177/1745691614528519

283. Pierce, J. (2019). CHI Conference on Human Factors in Computing Systems, 45, 1–14. https://doi.org/10.1145/3290605.3300275

284. Piercey, C. D., & Joordens, S. (2000). Turning an advantage into a disadvantage: Ambiguity effects in lexical decision versus reading tasks. Memory & Cognition, 28(4), 657–666. https://doi.org/10.3758/BF03201255

285. Planting, T., Koopowitz, S. M., & Stein, D. J. (2022). Coulrophobia: An investigation of clinical features. *The South African Journal of Psychiatry: SAJP: the Journal of the Society of Psychiatrists of South Africa*, *28*, 1653. https://doi.org/10.4102/sajpsychiatry.v28i0.1653

286. Pollak, S. D., Messner, M., Kistler, D. J., & Cohn, J. F. (2009). Development of perceptual expertise in emotion recognition. Cognition, 110(2), 242–247. https://doi.org/10.1016/j.cognition.2008.10.010

287. Proverbio, A. M., & Riva, F. (2009). RP and N400 ERP components reflect semantic violations in visual processing of human actions. Neuroscience letters, 459(3), 142–146. https://doi.org/10.1016/j.neulet.2009.05.012

288. Purcell, T. (1995). Experiencing American and Australian high- and popular-style houses. Environment and Behavior, 27(6), 771–800. https://doi.org/10.1177/0013916595276003

289. Purcell, A. T., & Nasar, J. L. (1992). Experiencing other people's houses: A model of similarities and differences in environmental experience. Journal of Environmental Psychology, 12(3), 199–211. https://doi.org/10.1016/S0272-4944(05)80135-5

290. Pütten, A. M. R.-v-d.., & Krämer, N. C. (2014). How design characteristics of robots determine evaluation and uncanny valley related responses. Computers in Human Behavior, 36, 422–439. https://doi.org/10.1016/j.chb.2014.03.066

291. Pyszczynski, T., Solomon, S., & Greenberg, J. (2015). Thirty Years of Terror Management Theory: From Genesis to Revelation, 52, 1–70. https://doi.org/10.1016/BS.AESP.2015.03.001

292. Ratajczyk, D., Dakowski, J., & Lupkowski, P. (2023). The importance of beliefs in human uniqueness for uncanny valley in virtual reality and on-screen. *International Journal of Human-Computer Interaction.* https://doi.org/10.1080/10447318.2023.2179216

293. Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience? Personality and Social Psychology Review, 8(4), 364–382. https://doi.org/10.1207/s15327957pspr0804_3

294. Reed, C. L., Stone, V. E., Bozova, S., & Tanaka, J. (2003). The body-inversion effect. Psychological Science, 14(4), 302–308. https://doi.org/10.1111/1467-9280.14431

295. Rémy, F., Saint-Aubert, L., Bacon-Macé, N., Vayssière, N., Barbeau, E., & Fabre-Thorpe, M. (2013). Object recognition in congruent and incongruent natural scenes: a life-span study. Vision research, 91, 36–44. https://doi.org/10.1016/j.visres.2013.07.006

296. Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic processing predicts face recognition. Psychological Science, 22(4), 464–471. https://doi.org/10.1177/0956797611401753

297. Rhodes, G., Brake, S., Taylor, K., & Tan, S. (1989). Expertise and configural coding in face recognition. British Journal of Psychology, 80(3), 313–331. https://doi.org/10.1111/j.2044-8295.1989.tb02323.x

298. Rhodes, G., Hayward, W. G., & Winkler, C. (2006). Expert face coding: Configural and component coding of own-race and other-race faces. Psychonomic Bulletin & Review, 13(3), 499–505. https://doi.org/10.3758/BF03193876

299. Rhodes, G., & Zebrowitz, L. A. (Eds.). (2002). Facial attractiveness: Evolutionary, cognitive, and social perspectives. Ablex Publishing.

300. Roesch, E. B., Tamarit, L., Reveret, L. Grandjean, D., Sander, D., & Scherer, K. R. (2010). FACSGen: A tool to synthesize emotional facial expressions through systematic manipulation of facial action units. *Journal of Nonverbal Behavior*, *35*, 1–16. https://doi.org/10.1007/s10919-010-0095-9

301. Romportl, J. (2014). Speech Synthesis and Uncanny Valley. In: Sojka, P., Horák, A., Kopeček, I., Pala, K. (eds) Text, Speech and Dialogue. TSD 2014. Lecture Notes in Computer Science(), vol 8655. Springer, Cham. https://doi.org/10.1007/978-3-319-10816-2_72

302. Rosenthal-von der Pütten, A. M., Krämer, N. C., Maderwald, S., Brand, M., & Grabenhorst, F. (2019). Neural Mechanisms for Accepting and Rejecting Artificial Social Partners in the Uncanny Valley. The Journal of neuroscience : the official journal of the Society for Neuroscience, 39(33), 6555–6570. https://doi.org/10.1523/JNEUROSCI.2956-18.2019

303. Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. Psychonomic Bulletin & Review, 21(2), 301–308. https://doi.org/10.3758/s13423-014-0595-4

304. Royle, N. (2003). The Uncanny. Routledge.

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

306. Ryali, C. K., Goffin, S., Winkielman, P., & Yu, A. J. (2020). From likely to likable: The role of statistical typicality in human social assessment of faces. Proceedings of the National Academy of Sciences of the United States of America, 117(47), 29371–29380. https://doi.org/10.1073/pnas.1912343117

307. Sacino, A., Cocchella, F., De Vita, G., Bracco, F., Rea, F., Sciutti, A., & Andrighetto, L. (2022). Human- or object-like? Cognitive anthropomorphism of humanoid robots. PloS one, 17(7), e0270787. https://doi.org/10.1371/journal.pone.0270787

308. Said, C. P., Dotsch, R., & Todorov, A. (2010). The amygdala and FFA track both social and non-social face dimensions. Neuropsychologia, 48(12), 3596–3605. https://doi.org/10.1016/j.neuropsychologia.2010.08.009

309. Salehi, E., Mehrabi, M., Fatehi, F., & Salehi, A. (2020). Virtual Reality Therapy for Social Phobia: A Scoping Review. Studies in health technology and informatics, 270, 713–717. https://doi.org/10.3233/SHTI200253

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

311. Santos, I. M., & Young, A. W. (2008). Effects of inversion and negation on social inferences from faces. Perception, 37(7), 1061–1078. https://doi.org/10.1068/p5278

312. Sasaki, K., Ihaya, K., & Yamada, Y. (2017). Avoidance of novelty contributes to the uncanny valley. Frontiers in Psychology, 8, Article 1792. https://doi.org/10.3389/fpsyg.2017.01792

313. Sato, W., Namba, S., Yang, D., Nishida, S., Ishi, C., & Minato, T. (2022). An Android for Emotional Interaction: Spatiotemporal Validation of Its Facial

Expressions. Frontiers in psychology, 12, 800657. https://doi.org/10.3389/fpsyg.2021.800657

314. Sato, W., Krumhuber, E. G., Jellema, T., & Williams, J. H. G. (2019). Editorial: Dynamic Emotional Communication. Frontiers in psychology, 10, 2836. https://doi.org/10.3389/fpsyg.2019.02836

315. Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience*, *7*(4), 413–422. https://doi.org/10.1093/scan/nsr025

316. Schaller, M., & Duncan, L. A. (2007). The behavioral immune system: Its evolution and social psychological implications. In J. P. Forgas, M. G. Haselton, & W. von Hippel (Eds.), Evolution and the social mind: Evolutionary psychology and social cognition (pp. 293–307). Routledge/Taylor & Francis Group.

317. Scherer, A., Von Wangenheim, F. (2014). Service with a smile or screen? How replacing personnel with machines affects customers' satisfaction with a service. *Advances in Consumer Research, 42*, 662–3.

318. Schindler, S., Zell, E., Botsch, M., & Kissler, J. (2017). Differential effects of face-realism and emotion on event-related brain potentials and their implications for the uncanny valley theory. Scientific reports, 7, 45003. https://doi.org/10.1038/srep45003

319. Schroeder, S. R., Rembrandt, H. N., May, S., & Freeman, M. R. (2020). Does having a voice disorder hurt credibility?. Journal of communication disorders, 87, 106035. https://doi.org/10.1016/j.jcomdis.2020.106035

320. Schroeder, S., Goad, K., Rothner, N., Momen, A., & Wiese, E. (2021). Effect of Individual Differences in Fear and Anxiety on Face Perception of Human and

Android Agents. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 65(1), 796–800. https://doi.org/10.1177/1071181321651303

321. Schweinberger, S. R., Kawahara, H., Simpson, A. P., Skuk, V. G., & Zäske, R. (2014). Speaker perception. Wiley interdisciplinary reviews. Cognitive science, 5(1), 15–25. https://doi.org/10.1002/wcs.1261

322. Schwind, V., Leicht, K., Jäger, S., Wolf, K., & Henze, N. (2018). Is there an uncanny valley of virtual animals? A quantitative and qualitative investigation. International Journal of Human-Computer Studies, 111, 49–61. https://doi.org/10.1016/j.ijhcs.2017.11.003

323. Searcy, J. H., & Bartlett, J. C. (1996). Inversion and processing of component and spatial-relational information in faces. Journal of experimental psychology. Human perception and performance, 22(4), 904–915. https://doi.org/10.1037//0096-1523.22.4.904

324. Seyama, J., & Nagayama, R.S. (2007). The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces. PRESENCE: Teleoperators and Virtual Environments, 16, 337-351. https://doi.org/10.1162/pres.16.4.337

325. Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. Psychological Bulletin, 134(2), 207–222. https://doi.org/10.1037/0033-2909.134.2.207

326. Shibata, H., Gyoba, J., & Suzuki, Y. (2009). Event-related potentials during the evaluation of the appropriateness of cooperative actions. Neuroscience letters, 452(2), 189–193. https://doi.org/10.1016/j.neulet.2009.01.042

327. Shklovski, I., Mainwaring, S. D., Skúladóttir, H. H., & Borgthorsson, H. (2014). Leakiness and creepiness in app space: Perceptions of privacy and mobile app

use. *CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 2347–2356. https://doi.org/10.1145/2556288.2557421

328. Shir, Y., Abudarham, N., & Mudrik, L. (2021). You won't believe what this guy is doing with the potato: The ObjAct stimulus-set depicting human actions on congruent and incongruent objects. Behavior research methods, 53(5), 1895–1909. https://doi.org/10.3758/s13428-021-01540-6

329. Sierra Rativa, A, Postma, M, van Zaanen, M. (2022). The uncanny valley of a virtual animal. *Computer Animation and Virtual Worlds*, *33*(2), e2043. https://doi.org/10.1002/cav.2043

330. Skiba, R. M., & Vuilleumier, P. (2020). Brain networks processing temporal information in dynamic facial expressions. Cerebral Cortex, 30(11), 6021–6038. https://doi.org/10.1093/cercor/bhaa176

331. Smarr, C., Mitzner, T. L., Beer, J. M., Prakash, A., Chen, T. Y., Kemp, C. C., & Rogers, W. A. (2014). Domestic Robots for Older Adults: Attitudes, Preferences, and Potential. International Journal of Social Robotics, 6(2), 229–247. https://doi.org/10.1007/s12369-013-0220-0

332. Sorokowski, P., Kościński, K., & Sorokowska, A. (2013). Is beauty in the eye of the beholder but ugliness culturally universal? Facial preferences of Polish and Yali (Papua) people. Evolutionary Psychology, 11(4), 907–925. https://doi.org/10.1177/147470491301100414

333. Stamps, A. E. III. (2007). Mystery of Environmental Mystery: Effects of Light, Occlusion, and Depth of View. Environment and Behavior, 39(2), 165–197. https://doi.org/10.1177/0013916506288053

334. Stamps, A. E. III, & Nasar, J. L. (1997). Design review and public preferences: Effects of geographical location, public consensus, sensation seeking,

and architectural styles. Journal of Environmental Psychology, 17(1), 11–32. https://doi.org/10.1006/jevp.1996.0036

335. Stansbury, L. G., Hess, A. S., Thompson, K., Kramer, B., Scalea, T. M., & Hess, J. R. (2013). The clinical significance of platelet counts in the first 24 hours after severe injury. Transfusion, 53(4), 783–789. https://doi.org/10.1111/j.1537-2995.2012.03828.x

336. Stein, J.-P., & Ohler, P. (2017). Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. Cognition, 160, 43–50. https://doi.org/10.1016/j.cognition.2016.12.010

337. Stekelenburg, J. J., & de Gelder, B. (2004). The neural correlates of perceiving human bodies: an ERP study on the body-inversion effect. Neuroreport, 15(5), 777–780. https://doi.org/10.1097/00001756-200404090-00007

338. Stone, A. (2021). Facial disfigurement, categorical perception, and the influence of Disgust Sensitivity. Visual Cognition, 29(2), 73–90. https://doi.org/10.1080/13506285.2020.1870184

339. Stone, A. (2022). Negative perceptions of people with facial disfigurement depend on a general attitude rather than on specific concerns. Stigma and Health, 7(3), 270–279. https://doi.org/10.1037/sah0000396

340. Strait, M.K., Vujovic, L., Floerke, V., Scheutz, M., & Urry, H.L. (2015). Too Much Humanness for Human-Robot Interaction: Exposure to Highly Humanlike Robots Elicits Aversive Responding in Observers. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. https://doi.org/10.1145/2702123.2702415

341. Strait, M. K., Floerke, V. A., Ju, W., Maddox, K., Remedios, J. D., Jung, M. F., & Urry, H. L. (2017). Understanding the Uncanny: Both Atypical Features and Category Ambiguity Provoke Aversion toward Humanlike Robots. Frontiers in psychology, 8, 1366. https://doi.org/10.3389/fpsyg.2017.01366

342. Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Uderspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*, 201–216. https://doi.org/10.3758/MC.36.1.201

343. Szczurek, L., Monin, B., & Gross, J. J. (2012). The Stranger effect: The rejection of affective deviants. Psychological Science, 23(10), 1105–1111. https://doi.org/10.1177/0956797612445314

344. Taylor, T. L., Hawton, K., Fortune, S., & Kapur, N. (2009). Attitudes towards clinical services among people who self-harm: systematic review. The British journal of psychiatry : the journal of mental science, 194(2), 104–110. https://doi.org/10.1192/bjp.bp.107.046425

345. Tene, O., & Polonetsky, J. (2013). A Theory of Creepy: Technology, Privacy and Shifting Social Norms. Yale Journal of Law and Technology, 16, 2.

346. Theeuwes, J., & Van der Stigchel, S. (2006). Faces capture attention: Evidence from inhibition of return. Visual Cognition, 13(6), 657–665. https://doi.org/10.1080/13506280500410949

347. Thompson, P. (1980). Margaret Thatcher: A new illusion. Perception, 9(4), 483–484. https://doi.org/10.1068/p090483

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

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

350. Tobin, A., Favelle, S., & Palermo, R. (2016). Dynamic facial expressions are processed holistically, but not more holistically than static facial expressions. Cognition and Emotion, 30(6), 1208–1221. https://doi.org/10.1080/02699931.2015.1049936

351. Tondu, B., & Bardou, N. (2011). A new interpretation of Mori's uncanny valley for future humanoid robots. *International Journal of Robotics and Automation*. http://dx.doi.org/10.2316/Journal.206.2011.3.206-3348

352. Torkamaan, H., Barbu, C., & Ziegler, J. (2019). How can they know that? A study of factors affecting the creepiness of recommendations. Proceedings of the 13th ACM Conference on Recommender Systems. https://doi.org/10.1145/3298689.3346982

353. Tu, Y. C., Chien, S. E., & Yeh, S. L. (2020). Age-Related Differences in the Uncanny Valley Effect. Gerontology, 66(4), 382–392. https://doi.org/10.1159/000507812

354. Tyson, P. J., Davies, S. K., Scorey, S., & Greville, W. J. (2023). Fear of clowns: An investigation into the aetiology of coulrophobia. *Frontiers in Psychology*, *14*, 1109466. https://doi.org/10.3389/fpsyg.2023.1109466

355. Tyson, P. J., Davies, S. K., Scorey, S., & Greville, W. J. (2022). Fear of clowns: An investigation into the aetiology of coulrophobia. *Frontiers in Psychology*, *14*, 1109466. https://doi.org/10.3389/fpsyg.2023.1109466

356. Uran, C., Peter, A., Lazar, A., Barnes, W., Klon-Lipok, J., Shapcott, K. A., Roese, R., Fries, P., Singer, W., & Vinck, M. (2022). Predictive coding of natural

images by V1 firing rates and rhythmic synchronization. Neuron, 110(7), 1240–1257.e8. https://doi.org/10.1016/j.neuron.2022.01.002

357. Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 43A(2), 161–204. https://doi.org/10.1080/14640749108400966

358. Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. Quarterly journal of experimental psychology (2006), 69(10), 1996–2019. https://doi.org/10.1080/17470218.2014.990392

359. Van de Cruys, S., Chamberlain, R., & Wagemans, J. (2017). Tuning in to art: A predictive processing account of negative emotion in art. *The Behavioral and brain sciences*, *40*, e377. https://doi.org/10.1017/S0140525X17001868

360. van de Riet, W. A., Grezes, J., & de Gelder, B. (2009). Specific and common brain regions involved in the perception of faces and bodies and the representation of their emotional expressions. Social neuroscience, 4(2), 101–120. https://doi.org/10.1080/17470910701865367

361. van Venrooij, L. T., & Barnhoorn, P. C. (2017). Coulrophobia: how irrational is fear of clowns?. *European Journal of Pediatrics*, *176*(5), 677. https://doi.org/10.1007/s00431-017-2896-x

362. Vartanian, O., Navarrete, G., Palumbo, L., & Chatterjee, A. (2021). Individual differences in preference for architectural interiors. Journal of Environmental Psychology, 77, Article 101668. https://doi.org/10.1016/j.jenvp.2021.101668

363. Vogel, D., Meyer, M., & Harendza, S. (2018). Verbal and non-verbal communication skills including empathy during history taking of undergraduate

medical students. BMC medical education, 18(1), 157. https://doi.org/10.1186/s12909-018-1260-9

364. Vogel T., Silva R. R., Thomas A., Wänke M. (2020). Truth is in the mind, but beauty is in the eye: Fluency effects are moderated by a match between fluency source and judgment dimension. Journal of Experimental Psychology: General, 149(8), 1587. https://doi.org/10.1037/xge0000731Wagenmakers et al., 2019

365. Walden P. R. (2022). Perceptual Voice Qualities Database (PVQD): Database Characteristics. Journal of voice : official journal of the Voice Foundation, 36(6), 875.e15–875.e23. https://doi.org/10.1016/j.jvoice.2020.10.001

366. Wang, S., Cheong, Y. F., Dilks, D. D., & Rochat, P. (2020). The Uncanny Valley Phenomenon and the Temporal Dynamics of Face Animacy Perception. Perception, 49(10), 1069–1089. https://doi.org/10.1177/0301006620952611

367. Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. Review of General Psychology, 19(4), 393–407. https://doi.org/10.1037/gpr0000056

368. Wänke, M., & Hansen, J. (2015). Relative processing fluency. Current Directions in Psychological Science, 24(3), 195–199. https://doi.org/10.1177/0963721414561766

369. Watt, M. C., Maitland, R. A., & Gallagher, C. E. (2017). A case of the "heeby jeebies": An examination of intuitive judgements of "creepiness". Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement, 49(1), 58–69. https://doi.org/10.1037/cbs0000066

370. Weinberger, A. B., Christensen, A. P., Coburn, A., & Chatterjee, A. (2021). Psychological responses to buildings and natural landscapes. Journal of

Environmental Psychology, 77, Article 101676. https://doi.org/10.1016/j.jenvp.2021.101676

371. Weis, P. P., & Wiese, E. (2017). Cognitive Conflict as Possible Origin of the Uncanny Valley. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61(1), 1599–1603. https://doi.org/10.1177/1541931213601763

372. Weisbuch, M., Ambady, N., Clarke, A. L., Achor, S., & Weele, J. V.-V. (2010). On being consistent: The role of verbal-nonverbal consistency in first impressions. Basic and Applied Social Psychology, 32(3), 261–268. https://doi.org/10.1080/01973533.2010.495659

373. Weisman, W. D., & Peña, J. F. (2021). Face the uncanny: The effects of doppelganger talking head avatars on affect-based trust toward artificial intelligence technology are mediated by uncanny valley perceptions. Cyberpsychology, Behavior, and Social Networking, 24(3), 182–187. https://doi.org/10.1089/cyber.2020.0175

374. Westfall, J., Nichols, T. E., & Yarkoni, T. (2016). Fixing the stimulus-as-fixed-effect fallacy in task fMRI. Wellcome open research, 1, 23. https://doi.org/10.12688/wellcomeopenres.10298.2

375. Whittlesea, B. W. A., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? A discrepancy-attribution account of feelings of familiarity. Acta Psychologica, 98(2-3), 141–165. https://doi.org/10.1016/S0001-6918(97)00040-1

376. Wiener, J. M., Büchner, S. J., & Hölscher, C. (2009). Taxonomy of human wayfinding tasks: A knowledge-based approach. Spatial Cognition and Computation, 9(2), 152–165. https://doi.org/10.1080/13875860902906496

377. Wiese, H., & Schweinberger, S. R. (2008). Event-related potentials indicate different processes to mediate categorical and associative priming in person

recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(5), 1246–1263. https://doi.org/10.1037/a0012937

378. Widmayer, S. A. (2002). Schema Theory: An Introduction', George Mason University Instructional Technology Program. Consulted on, 16.

379. Category:Liminal spaces. (2021) Retrieved from https://commons.wikimedia.org/wiki/Category:Liminal_spaces, Accessed 31st May 2023.

380. Wilson, M. C., & Scior, K. (2014). Attitudes towards individuals with disabilities as measured by the implicit association test: a literature review. Research in developmental disabilities, 35(2), 294–321. https://doi.org/10.1016/j.ridd.2013.11.003

381. Winkielman, P., Schwarz, N., Fazendeiro, T. A., & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch & K. C. Klauer (Eds.), The psychology of evaluation: Affective processes in cognition and emotion (pp. 189–217). Lawrence Erlbaum Associates Publishers.

382. Winkielman, P., Halberstadt, J., Fazendeiro, T., & Catty, S. (2006). Prototypes are attractive because they are easy on the mind. Psychological science, 17(9), 799–806. https://doi.org/10.1111/j.1467-9280.2006.01785.x

383. Winkielman, P., Olszanowski, M., & Gola, M. (2015). Faces in-between: Evaluations reflect the interplay of facial features and task-dependent fluency. Emotion, 15(2), 232–242. https://doi.org/10.1037/emo0000036

384. Wong, Y. K., Twedt, E., Sheinberg, D., & Gauthier, I. (2010). Does Thompson's Thatcher Effect reflect a face-specific mechanism?. Perception, 39(8), 1125–1141. https://doi.org/10.1068/p6659

385. Wong, A. C. -N., Wong, Y. K., Lui, K. F. H., Ng, T. Y. K., & Ngan, V. S. H. (2019). Sensitivity to configural information and expertise in visual word recognition. Journal of Experimental Psychology: Human Perception and Performance, 45(1), 82–99. https://doi.org/10.1037/xhp0000590

386. Workman, A. M., Heaton, M. P., Webster, D. A., Harhay, G. P., Kalbfleisch, T. S., Smith, T. P. L., Falkenberg, S. M., Carlson, D. F., & Sonstegard, T. S. (2021). Evaluating Large Spontaneous Deletions in a Bovine Cell Line Selected for Bovine Viral Diarrhea Virus Resistance. Viruses, 13(11), 2147. https://doi.org/10.3390/v13112147

387. Yam, K. C., Bigman, Y., & Gray, K. (2021). Reducing the uncanny valley by dehumanizing humanoid robots. Computers in Human Behavior, 125, Article 106945. https://doi.org/10.1016/j.chb.2021.106945

388. Yamada, Y., Kawabe, T., & Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. Japanese Psychological Research, 55(1), 20–32. https://doi.org/10.1111/j.1468-5884.2012.00538.x

389. Yap, M. J., Tan, S. E., Pexman, P. M., & Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. Psychonomic bulletin & review, 18(4), 742–750. https://doi.org/10.3758/s13423-011-0092-y

390. Yarkoni, T. (2022). The generalizability crisis. Behavioral and Brain Sciences, 45, Article e1. https://doi.org/10.1017/S0140525X20001685

391. Yerkes R. M., Dodson J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology. 18*(5), 459–482. https://doi.org/10.1002/cne.920180503

392. Yin, J., Wang, S., Guo, W., & Shao, M. (2021). More than appearance: The uncanny valley effect changes with a robot's mental capacity. Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues. Advance online publication. https://doi.org/10.1007/s12144-021-02298-y

393. Young, A. W., Frühholz, S., & Schweinberger, S. R. (2020). Face and voice perception: Understanding commonalities and differences. Trends in Cognitive Sciences, 24(5), 398–410. https://doi.org/10.1016/j.tics.2020.02.001

394. Zajonc, R. B. (1968). Attitudinal effects of mere exposure. Journal of Personality and Social Psychology, 9(2, Pt.2), 1–27. https://doi.org/10.1037/h0025848

395. Zebrowitz, L. A., & Rhodes, G. (2004). Sensitivity to "Bad Genes" and the Anomalous Face Overgeneralization Effect: Cue Validity, Cue Utilization, and Accuracy in Judging Intelligence and Health. Journal of Nonverbal Behavior, 28(3), 167–185. https://doi.org/10.1023/B:JONB.0000039648.30935.1b

396. Zhang, B., & Xu, H. (2016). Privacy Nudges for Mobile Applications: Effects on the Creepiness Emotion and Privacy Attitudes. Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. https://doi.org/10.1145/2818048.2820073

397. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2017). Places_ A 10 Million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(6), 1452–1464.

398. Zlotowski, J., Bartneck, C. (2013) The inversion effect in HRI: are robots perceived more like humans or objects? Tokyo, Japan: 8th ACM/IEEE international conference on Human-robot interaction, 3-6-Mar 2013. Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction, 365-372.

399. Złotowski, J. A., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C., & Ishiguro, H. (2015). Persistence of the uncanny valley: the influence of repeated interactions and a robot's attitude on its perception. Frontiers in psychology, 6, 883. https://doi.org/10.3389/fpsyg.2015.00883

Appendix

Figure A1

Distortion procedure visualized on a single greeble (Chapter 3). The same procedure was used for every distorted variant.

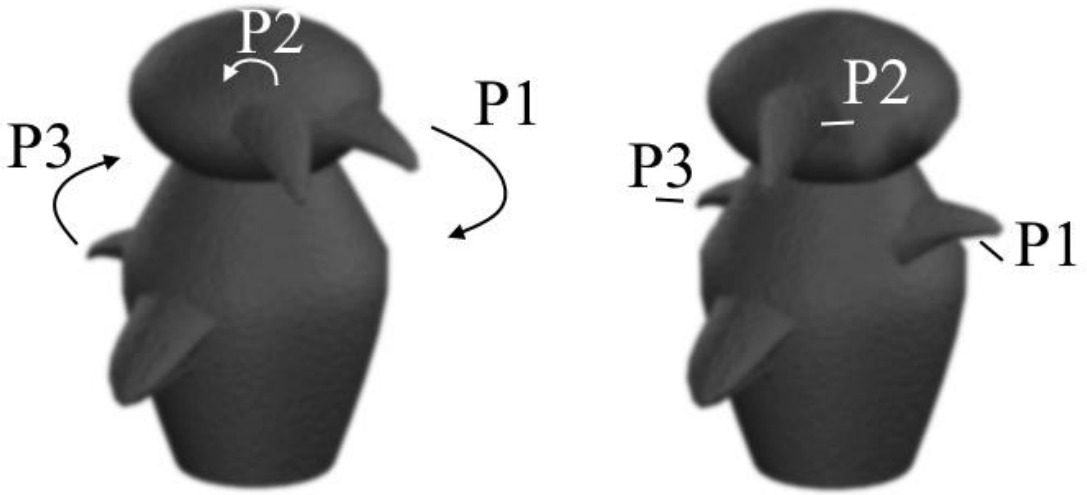


Figure A2

One distorted greeble per family (upper row) and its normal variant (lower row) (Chapter 3).

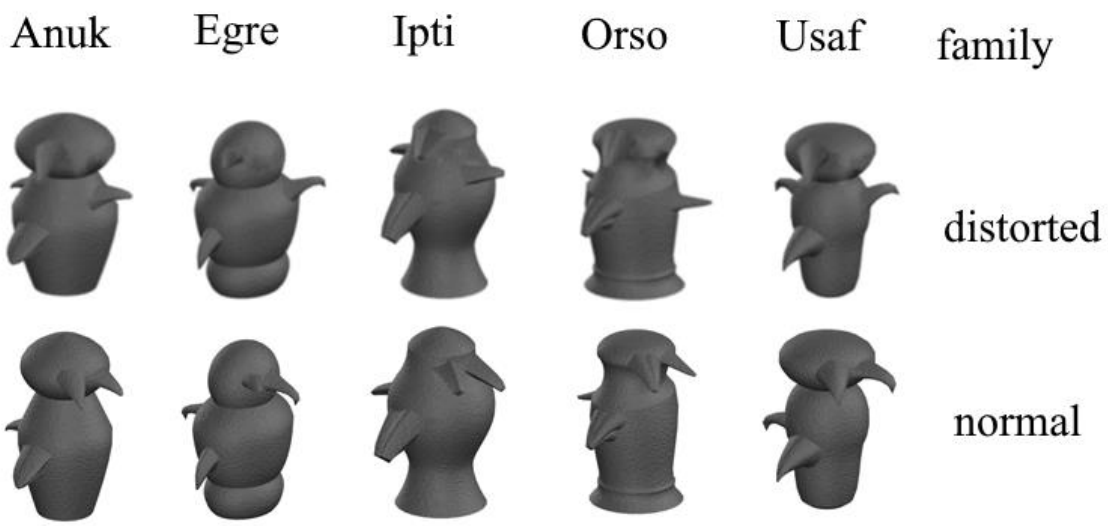


Figure A3

Morphing landmarks for the total average morph in Fantamorph Deluxe (Chapter 3). Pairs of greebles were morphed together, here the morphed averages of family 1 (left) and 2 (right). Afterwards, the result was morphed with the morph between the averages of family 3 and 4, and finally with the average of family 5 with an 80:20 weighting to create a total average. After each morphing procedure morph noise was cleaned using Photoshop CS6.

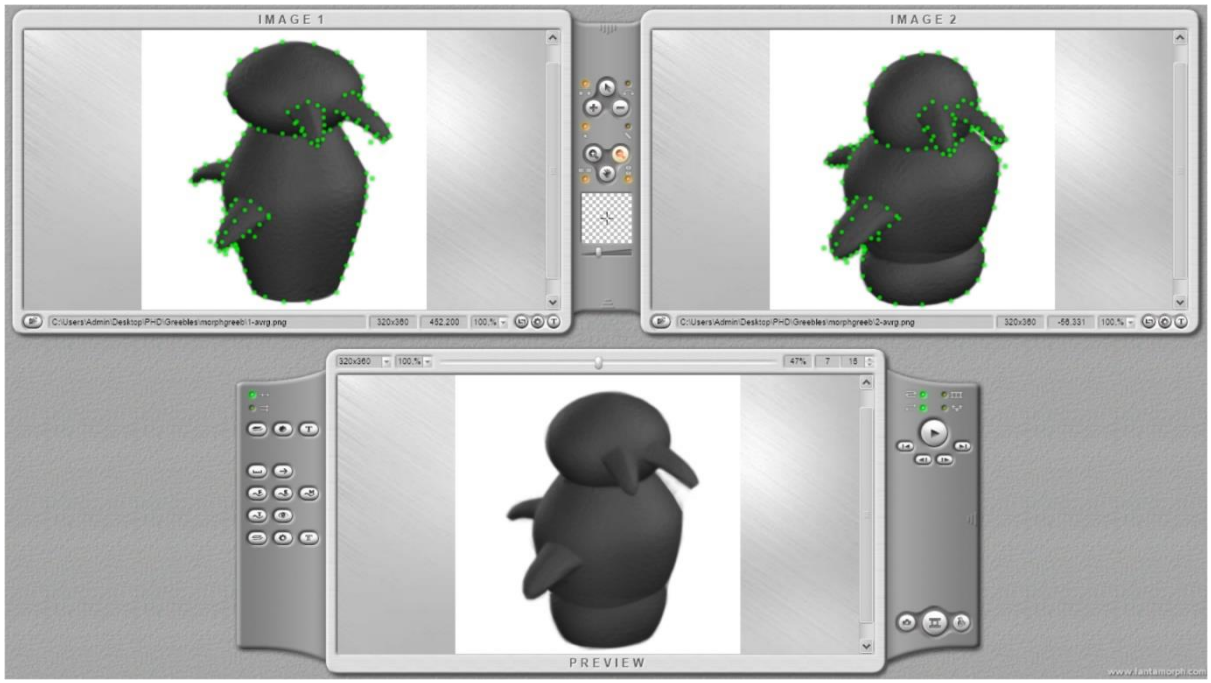


Table A1

Experiment procedure for both control and training group (Chapter 3). Numbers represent number of trials and numbers in brackets represent number of individual greebles the participant has been introduced to before in an "individual viewing" task, and a number in brackets plus "new" indicates the number of new greebles shown. "Rating" refers to either the control or post-training rating session.

Procedure (number of individual greebles shown)	Session 1	Session 2	Session 3	Session 4	Rating
Family examples (10)	1				

Family viewing (25)	25				
Family naming (30)	30				
Individual viewing (5)	20 (5)				
Individual naming with feedback (5)	15 (5)				
Individual naming (30)	60 (5)				
Verification (30)	125 (5)				
Family naming (30)	30				
Individual viewing (previously learned)	10 (5)	20 (5)	40 (10)	60 (15)	
Individual naming (30)	60 (5)				
Verification (30)	125 (5)	125 (5)	130 (10)	120 (15)	
Individual viewing (5)		20 (5 new)	20 (5 new)	20 (5 new)	
Individual naming with feedback (previously learned)		30 (10)	45 (15)	60 (20)	
Individual naming (30)			60 (15)	60 (20)	
Verification (30)		130 (10)	125 (15)	120 (20)	
Individual naming (30)		60 (10)	60 (15)	60 (20)	
Verification (30)		130 (10)	125 (15)	120 (20)	
Individual naming (30)		60 (10)	60 (15)	60 (20)	
Final verficiation				120 (20)	
Rating task (41)					41

Table A2

Unedited English and Icelandic sentences used in the first part of Chapter 4.

English	Icelandic

In those days, those distant days.	Á þessum dögum, á þessum fjarlægju dögum.
He lives outside the city.	Hann býr fyrir utan borgina.
There was a single tree.	Það var eitt tré.
His intuition led him to the forest.	Innsæið leiddi hann inn í skóginn.
He eats bread.	Hann borðar brauð.
They hugged and kissed.	Þau knúsuðust og kisstust.
They hit him and struck him.	Þeir slógu og börðu hann.
The king left the city.	Kóngurinn er farinn úr borginni.
He sat down in the dust.	Hann settist niður í rykið.

Table A3

Target word stimuli and the context words used in the second part of Chapter 4.

Target word	Context words	
	Ambiguous condition	Non-ambiguous condition
Act	Behaviour, Theatre	Animal, Theatre
Cause	Reason, Goal	Food, Goal
Block	Material, Mental	Clothing, Mental
Key	Lock, Typewriter	Lock, Alcohol
Board	Surfing, Ironing	Surfing, Grammar
Company	Social, Liquid	Social, Liquid
Case	Police, Grammar	Animal, Police
Beam	Laser, Construction	Clothing, Construction
Class	School, Social	Food, School
Space	Public, Cosmic	Public, Weapon
Magazine	Gun, Paper	Paper, Building

Oil	Fuel, Cooking	Singing, Cooking
Article	Paper, Grammar	Electrical, Grammar
Vision	Physical, Political	Cooking, Sense
Film	Coating, Movie	Food, Movie

Table A4

Detailed information on voice stimuli used in Chapter 7, Experiment 1. Distorted voices are not listed as their values were identical to their typical voice counterparts.

Voice type	Stimulus	gender	Severity rating and diagnosis (pathological); Speaker/source (synthetic)	Duration (sec)
typical	1	Female		4
	2	Female		4
	3	Male		5
	4	Male		4
	5	Male		4
	6	Female		4
	7	Male		4
	8	Female		4
	9	Male		3
	10	Female		4
	11	Female		4
	12	Female		4
	13	Male		4

	14	Female		4
	15	Female		4
pathological	1	Female	98.67; Reinke's Edema	5
	2	Male	98.5; lesions	9
	3	Female	97.5; lesions	5
	4	Male	95.5; ulcerative laryngitis	4
	5	Male	89.17; Reinke's Edema	5
	6	Female	88.83; unilateral vocal fold paresis	5
	7	Female	88.17; atrophy, MTD	4
	8	Male	87.33; lesions	4
	9	Female	86.33; NA	8
	10	Male	86; vocal fold paresis	5
	11	Female	85.5; unilateral vocal fold paresis	4
	12	Female	85.33; MTD	7
	13	Female	83.83; unilateral vocal fold paresis	3
	14	Male	81.5; unilateral vocal fold paresis	5
	15	Female	78.17; Reinke's Edema	4
synthetic	1		eSpeak (Stephen Hawking voice generator)	4
	2	Male	Google	3
	3	NA	Mechanical sounds	5
	4	NA	Mechanical sounds	5
	5	Male	Microsoft Azure	3
	6	Female	Microsoft Azure	5
	7	Female	Microsoft Azure	5

8	Male	Microsoft Sam	4
9	NA	R2D2 sounds	5
10	NA	R2D2 sounds	5
11	Female	Watson	3
12	Female	Watson	3
13	Male	Watson	3
14	Female	Watson	3
15	Male	Watson	3

Table A5

Detailed information on voice stimuli used in Chapter 7, Experiment 2. Distorted voices are not listed as their values were identical to their typical voice counterparts.

Voice type	Stimulus	gender	Severity rating and diagnosis (pathological)	Duration (sec)
typical	1	Female		4
	2	Female		4
	3	Male		5
	4	Male		4
	5	Female		4
pathological	1	Female	23.83; muscle tension dysphonia	4
	2	Female	38.83; vocal fold paresis	4
	3	Male	74.33; vocal fold paresis	4
	4	Female	98.67; Reinke's Edema	4

5	Female	23.67; muscle tension dysphonia, atrophy	4
6	Female	48.17; muscle tension dysphonia	4
7	Female	74; lesions	4
8	Male	98.5; lesions	5
9	Female	22.67; paradoxical vocal fold movement	4
10	Female	47.17; adductor spasmic dysphonia	4
11	Female	73.83; NA	5
12	Female	97.5; lesions	5
13	Female	23.33; NA	4
14	Female	46.17; leucoplakia	4
15	Female	73.3; muscle tension dysphonia	5
16	Male	95.5; ulcerative laryngitis	4
17	Male	22.25; NA	4
18	Male	46.17; unilateral vocal fold paresis	4
19	Male	73.17; unilateral vocal fold paresis	7
20	Male	89.17; Reinke’s Edema	5