A meta-analysis of the uncanny valley's independent and dependent variables

Alexander Diel

School of Psychology, Cardiff University, 70 Park Place, Cardiff CF10 3AT, United Kingdom, diela@cardiff.ac.uk

Sarah Weigelt

Department of Vision, Visual Impairments & Blindness, Faculty of Rehabilitation Sciences, Technical University of Dortmund, Emil-Figge-Straße 50, 44227 Dortmund, Germany, sarah.weigelt@tu-dortmund.de

Karl F. MacDorman

School of Informatics and Computing, Indiana University, 535 West Michigan St., Indianapolis, IN 46202, USA, kmacdorm@indiana.edu

ABSTRACT

The uncanny valley (UV) effect is a negative affective reaction to human-looking artificial entities. It hinders comfortable, trust-based interactions with android robots and virtual characters. Despite extensive research, a consensus has not formed on its theoretical basis or methodologies. We conducted a meta-analysis to assess operationalizations of human likeness (independent variable) and the UV effect (dependent variable). Of 468 studies, 72 met the inclusion criteria. The studies employed 10 different stimulus creation techniques, 39 affect measures, and 14 indirect measures. Based on 247 effect sizes, a three-level meta-analysis model revealed the UV effect had a large effect size, Hedges' $g = 1.01$ [0.80, 1.22]. A mixed-effects meta-regression model with creation technique as the moderator variable revealed *face distortion* produced the largest effect size, $g = 1.46$ [0.69, 2.24], followed by *distinct entities*, $g = 1.20$ [1.02, 1.38], *realism render,* $g = 0.99$ [0.62, 1.36], and *morphing*, $g = 0.94$ [0.64, 1.24]. Affective indices producing the largest effects were *threatening, likable*, *aesthetics*, *familiarity*, and *eeriness*, and indirect measures were *dislike frequency*, *categorization reaction time*, *like frequency*, *avoidance*, and *viewing duration*. This meta-analysis—the first on the UV effect—provides a methodological foundation and design principles for future research.

CCS Concepts

• Human-centered computing → HCI design and evaluation methods; • Computer systems organization → External interfaces for robotics; • Computing methodologies → Animation

Keywords

Anthropomorphism, computer animation, face perception, robotics, uncanny valley

1 Introduction

Royle (2003) gives an evocative and succinct description of the uncanny experience:

> The uncanny is ghostly. It is concerned with the strange, weird, and mysterious, with a flickering sense (but not conviction) of something supernatural. The uncanny involves feelings of uncertainty, in particular regarding the reality of who one is and what is being experienced. (p. 1)

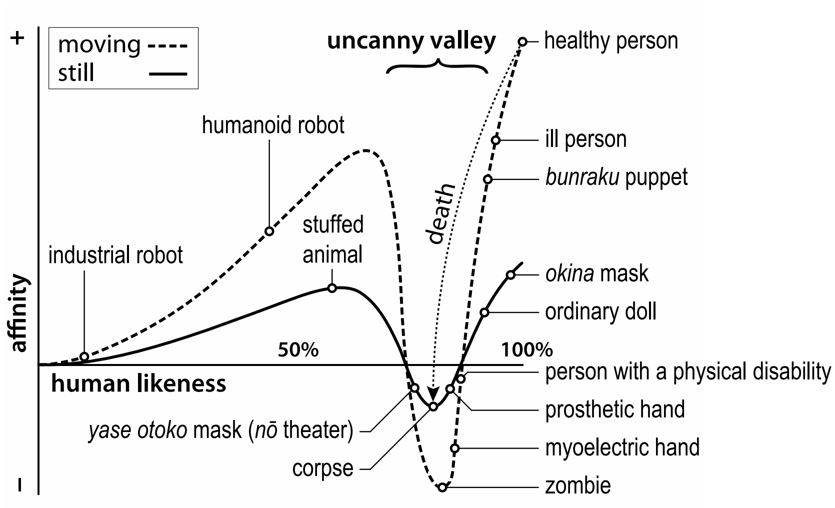

Figure 1. The uncanny valley as proposed by Mori in 1970. The affective reaction towards an entity (y-axis) is a function of its degree of human likeness (x-axis) and whether it is still or moving (solid or dashed line). Bunraku puppets play character roles in ningyō jōruri, a traditional form of musical puppet theater in Japan. Actors in nō theater wear masks: The yase otoko mask (literally, thin man) signifies a ghost from hell, and the okina mask signifies an old man.

Objects, situations, and events that do not fit our everyday understanding of the world are often described as eerie, creepy, or uncanny. These ascriptions can be made regarding new technologies (Langer & König, 2018), unusual human behavior (McAndrew & Koehnke, 2016), or peculiar coincidences (Freud, 1919/2003). Negative evaluations can hinder the adoption of supportive products like healthcare robots (Olaronke, Ojerinde, & Ikono, 2017) or service chatbots (Ciechanowski, Przegalińska, Magnuski, & Gloor, 2019). As the robotics pioneer Mori proposed in 1970, human-looking androids and other objects could elicit a reaction unlike the one typically elicited by people or stylish technology. Mori (2012) illustrated this phenomenon with a graph (Figure 1). The y-axis depicts *affinity*, the dependent variable (DV), as a function of *human likeness,* the independent variable (IV), on the x-axis (Bartneck, Kulić, Croft, & Zoghbi, 2009b; Ho & MacDorman, 2010, 2017; MacDorman & Ishiguro, 2006). The stimulus sets in Figure 2 show how different creation techniques have been used to operationalize the independent variable.

According to Mori (2012), affinity for an entity increases with its human likeness but only up to a point. Beyond this point, affinity falls and becomes negative, and the entity elicits a cold, eerie, repellant feeling. Then, affinity rises again, becoming positive, as human likeness increases toward indistinguishability. When graphed, the fall and rise in affinity resemble a valley—hence, the term *uncanny valley* (UV).

Since Mori's proposal, a substantial body of research has replicated a valley-shaped curve and found a significant effect (Burleigh, Schoenherr, & Lacroix, 2013; Ferrey, Burleigh, & Fenske, 2015; Jung & Cho, 2018; MacDorman, Green, Ho, & Koch, 2009; Mäkäräinen, Kätsyri, & Takala, 2014; Mathur & Reichling, 2016; Mathur et al., 2020; McDonnell, Breidt, & Bülthoff, 2012; Palomäki et al., 2018; Sasaki, Ihaya, & Yamada, 2017; Strait et al., 2017; Strait, Vujovic, Floerke, Scheutz, & Urry, 2015; Tinwell, Grimshaw, & Nabi, 2015; Tinwell, Grimshaw, Nabi, & Williams, 2011; Tinwell & Sloan, 2014; Yamada, Kawabe, & Ihaya, 2013). However, some studies have plotted functions other than a valley-shaped curve: For example, Kätsyri, de Gelder, and Takala (2019) found affinity increased with human likeness, an "uncanny slope"; Cheetham, Suter, and Jäncke (2014) interpreted increasing familiarity ratings with the transition from avatar to ambiguous morph to human as a "happy valley"; and Bartneck, Kanda, Ishiguro, and Hagita (2009a) and Cheetham, Wu, Pauli, and Jäncke (2015) found no difference in affective responses toward androids and humans. Although the UV effect is seldom disputed, its theoretical basis and methodologies have eluded consensus. This motivated us to examine how the independent and dependent variables in Mori's graph have been operationalized in the literature.

Although several reviews have examined the UV effect (Kätsyri, Förger, Mäkäräinen, & Takala, 2015; Lay, Brace, Pike, & Pollick, 2016; Wang, Lilienfeld, & Rochat, 2015; Zhang et al., 2020), this is the first meta-analysis to do so. It confirmed the effect's significance and determined its effect size. This is also, of course, the first meta-analysis to evaluate the uncanny valley's stimulus creation methods and affect and indirect measures. The evaluation was accomplished using meta-regression models. From the results, we distill design principles for future experiments.

The UV effect has been conceptualized in different ways. These conceptualizations often stem from different theories and their assumptions about elicitors of the effect (Diel & MacDorman, 2021). They include

1. a function like Mori's graph that maps a given degree of human likeness to a level of affect (Bartneck et al., 2009a; Burleigh, Schoenherr, & Lacroix, 2013; Chen, Russel, Nakayama, & Livingstone, 2010; Gray & Wegner, 2012; Kätsyri, de Gelder, & Takala, 2019; Lin et al, 2021; Ramey, 2005; Sasaki, Ihaya, & Yamada, 2017; Schneider, Wang, & Yang, 2009; Schwind et al., 2018; Seyama & Nagayama, 2007);
2. deviations from norms of human appearance and movement (Chaminade, Hodgins, & Kawato, 2007; MacDorman & Ishiguro, 2006; Mathur & Reichling, 2016; Palomäki et al., 2018; Schoenherr & Burleigh, 2015; Seyama & Nagayama, 2007; Tinwell, 2009; Tinwell & Grimshaw, 2009; Tinwell, Grimshaw, & Nabi, 2014);
3. violations of expectations of human appearance and behavior (Bartneck et al., 2009a; MacDorman & Ishiguro, 2006);
4. sensitivity to nonhuman features that increases with an entity's human likeness (Chattopadhyay & MacDorman, 2016; Green, MacDorman, Ho, & Vasudevan, 2008; MacDorman, Srinivas, & Patel, 2013);

5. a mismatch between human and nonhuman features (Ho & MacDorman, 2010; MacDorman, Green, Koch, & Ho, 2009; Mitchell et al., 2011b; Moore, 2012; Takahashi, Fukuda, Samejima, Watanabe, & Ueda, 2015; Tinwell & Sloan, 2014);
6. entities that elicit the concept *human* but have nonhuman traits (Steckenfinger & Ghazanfar, 2009); and
7. difficulty distinguishing between categories, such as *human* and *robot*, or a conflict between categories (Cheetham, Pavlović, Jordan, Suter, & Jäncke, 2013; Cheetham, Suter, & Jäncke, 2011, 2014; Cheetham, Wu, Pauli, & Jäncke, 2015; Matsuda, Okamoto, Ida, Okanoya, & Myowa-Yamakoshi, 2012).

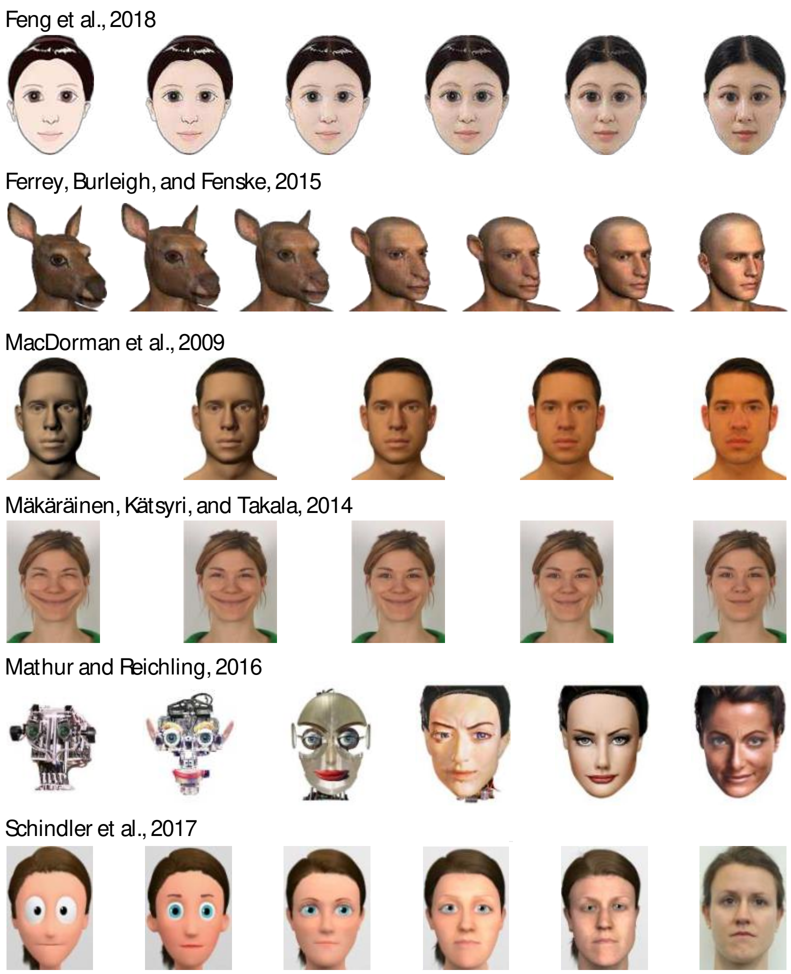

Figure 2. Different operationalizations of the independent variable **human likeness** ***(Feng et al., 2018; Ferrey, Burleigh, and Fenske, 2015; MacDorman et al., 2009; Mäkäräinen, Kätsyri, & Talaka, 2014, derived from Langner et al., 2010; Mathur & Reichling, 2016; Schindler et al., 2017).***

1.1 The independent variable

1.1.1 Construct

In experiments on the UV effect, the independent variable is typically *human likeness* or a similar term. However, it is unclear precisely how human likeness relates to the UV curve. Human likeness can be characterized along many dimensions, which interact to create an overall impression of humanness (Bartneck et al., 2009b; von Zitzewitz, Boesch, Wolf, & Riener, 2013). Mori (2012) examines both the outward appearance and the behavior of androids, corpses, and industrial and toy robots. In discussing mannequins, prostheses, and *bunraku* puppets, he draws in other dimensions, such as the setting, lighting, story, time of day, and the perceiver's gender and distance. Research corroborates the multidimensionality of human likeness in exploring the relation between the UV effect and an entity's physical (MacDorman & Ishiguro, 2006; Seyama & Nagayama, 2007), behavioral (MacDorman et al., 2005; Złotowski et al., 2015), and perceived mental similarity to humans (Gray & Wegner, 2012; Stein & Ohler, 2017). The perception of nonhuman animals can also elicit the UV effect (Chattopadhyay & MacDorman, 2016; Löffler, Dörenbächer, & Hassenzahl, 2020; Schwind et al., 2018; Takahashi et al., 2015; Yamada, Kawabe, & Ihaya, 2013). This result casts doubt on whether the independent variable solely concerns *human* likeness. *Realism* or *zoomorphism* have served as alternative concepts. Furthermore, Mori (2012) uses human likeness to denote interchangeably both an entity's physical properties and how it is perceived. In research, however, the distinction is necessary. Physical properties, for example, can be directly manipulated as an independent variable.

1.1.2 Stimulus range

We compiled a list of categories to summarize stimulus creation techniques. The list derives from the stimuli appearing in publications of empirical research and descriptions of how they were created (e.g., Mitchell et al., 2011b; Seyama & Nagayama, 2007). We started with six *a priori* categories and added categories during the literature search when a paper's stimuli did not fit in any existing category. Saturation was reached at 10 categories. The categories encompass the research reviewed, enabling its techniques to be easily classified, and reflect its theoretical and methodological breadth. The 10 categories of techniques are listed below:

Distinct entities: Selecting images or videos of existing robots, androids, computer-animated characters, humans, or other entities (e.g., Mathur et al., 2020). This technique is theory-independent and can be used with both still and moving entities, such as characters from films, video games, and virtual worlds.

Emotion manipulation: Distorting affective expressions (e.g., Qiao & Roger, 2011; Qiao, Eglin, & Beck, 2011; Tinwell et al., 2014). This technique visually manipulates the emotional expression of the face. It has been used mainly to test empathy-related theories.

Face distortion: Distorting facial features and proportions (e.g., Mäkäräinen et al., 2014). This technique visually manipulates facial features or the relations among them until the face no longer appears real. The emotional expression is not intentionally manipulated. This technique has been used to test theories related to configural processing (e.g., MacDorman et al., 2009).

Mismatch: Swapping facial features with those of another face that differs along one or more dimensions—typically animacy, human likeness, and realism (e.g., Seyama & Nagayama, 2007). This technique has been used to test theories related to perceptual mismatch (MacDorman & Chattopadhyay, 2016).

Morphing: Varying the stimulus in a stepwise transition between a pair of images to create a range of stimuli (e.g., MacDorman & Ishiguro, 2006). This technique has been used to transform the stimulus gradually from one kind of entity to another, thus making it suitable for testing category-related theories (e.g., Cheetham et al., 2015; Sasaki, Ihaya, & Yamada, 2017).

Motion manipulation: Distorting an animation's biological motion (e.g., gait, Destephe et al., 2014; Handzic & Reed, 2015; motion quality, Piwek, McKay, & Pollick, 2014; Thompson, Trafton, & McKnight, 2011). This technique has been used to test whether the UV effect occurs in motion perception.

Realism render: Varying how real the stimuli appear by representing them as cartoons or as computer models with a reduced polygon count or simplified textures (e.g., McDonnell et al., 2012; Muniady & Ali, 2020). This technique is theory-independent and relevant to practical applications of visual design.

Real-life encounter: Presenting different embodied entities like robots, androids, and humans for observation or interaction (e.g., Złotowski et al., 2015). This technique encompasses multiple modalities and, thus, can be used to measure a holistic UV effect. It is also useful because a physical object could be perceived and evaluated differently from its two-dimensional depiction (Snow, Skiba, Coleman, & Berryhill, 2014). Moreover, this technique is ecologically valid.

Visuo-auditory mismatch: Replacing a human voice with a synthesized voice or vice versa in an animation (e.g., Mitchell et al., 2011b; Stein & Ohler, 2018). Although typically motivated by perceptual mismatch theories, this technique differs from the *mismatch* category because the mismatch is crossmodal.

Voice distortion: Distorting natural human voices as auditory stimuli (e.g., Baird et al., 2018; Kühne et al., 2020). This technique has been used to test whether the UV effect can occur solely within audition.

1.1.3 Measurement

To assess the degree of human likeness (or related concepts), either single-scale measures or indices consisting of multiple scales have been used (e.g., Burleigh, Schoenherr, & Lacroix, 2013; Ho & MacDorman, 2010, 2017). Experiments typically vary the stimulus systematically in its degree of human similarity. Manipulations include distorting it (Mäkäräinen, Kätsyri, & Takala, 2014) or controlling its morphing proportion between two images (Cheetham & Jäncke, 2013). Experiments may include a manipulation check, such as rating the stimulus on human likeness. For computer-modeled stimuli only, Burleigh, Schoenherr, and Lacroix (2013) proposed two objective properties, which they define as follows: texture resolution, the number of pixels per unit of surface area, and polygon count, the number of polygons constituting a three-dimensional model. However, human likeness and realism are two different constructs. Thus, the results of a study measuring human likeness may not be comparable to the results of a study measuring realism. Research has not compared how changes in these independent variables or others may influence affect measures differently.

1.2 The dependent variable

1.2.1 Construct

Mori (2012) represents the *y*-axis with the term *shinwakan*, a neologism he translates as *affinity*. The *y*-axis had initially been translated as *familiarity* (Reichardt, 1978). Other proposed constructs include *interpersonal warmth* (or *likability*) and reverse-scaled *eeriness* (Bartneck et al., 2009b; Ho & MacDorman, 2010, 2017; Redstone, 2013). *Eeriness* and its synonym *creepiness* correlate with aversive experiences like disgust, fear, and anxiety (Ho, MacDorman, & Pramono, 2008).

1.2.2 Measurement

In experiments on the UV effect, the dependent variable is typically measured with single-scale measures or indices composed of self-reported affective items. Semantic differential scales are common. Semantically, some items like *eerie, creepy,* and *uncanny* are specific and, on face value, capture the distinctive experiential quality of the UV effect (Ho & MacDorman, 2010; Mangan, 2015; Palomäki et al., 2018; Redstone, 2013; Tinwell, Nabi, & Charlton, 2013). Other items like *pleasantness* or *likability* are *nonspecific*. An entity could rate low on them without being uncanny at all (e.g., items in Bartneck et al., 2009b; Ferrey, Burleigh, & Fenske, 2015; Rosenthal–von der Pütten & Krämer, 2014; Yamada, Kawabe, & Ihaya, 2013).

Questionnaires that have been developed to evaluate robots in general have been repurposed to measure the UV effect. Examples include the Godspeed indices (Bartneck et al., 2009b) and the Robotic Social Attribution Scale (Carpinella, Wyman, Perez, & Stroessner, 2017). Ho and MacDorman's (2010, 2017) set of indices includes *humanness, interpersonal warmth, attractiveness,* and *eeriness*. They developed the set to decorrelate these dimensions so they could be plotted against each other on orthogonal axes.

Indirect measures may indicate a construct by measuring a different construct. For example, the UV effect may correlate with trust behavior (Mathur & Reichling, 2016). For simplicity, we categorize implicit measures as indirect measures. Implicit measures center on processes that are automatic, effortless, fast, goal-independent, stimulus-driven, uncontrolled, or unintentional. For example, response time and other performance measures of the UV effect typically are implicit measures. Implicit measures counter self-presentational bias, that is, respondents' attempts to influence how others perceive them. Implicit measures may indicate the UV effect in otherwise inaccessible populations, such as infants or nonhuman animals.

Apart from trust behavior, the UV effect has been measured by such indirect measures as avoidance behavior (Matsuda et al., 2012), perceived responsiveness (Tinwell et al., 2013), and cognitive conflict and categorization reaction time (RT, Cheetham & Jäncke, 2013).

1.2.3 Other constructs

Other constructs and their associated measures and theories include the following:

Aesthetics: Items measuring aesthetic appeal (Sansoni, Wodehouse, McFayden, & Buis, 2015; Schwind et al., 2018). These items conceptualize the UV effect as a lack of physical attractiveness. Thus, they can serve as a practical tool for design (Hanson et al., 2005; Ho & MacDorman, 2010, 2017). Research has used nonhuman (e.g., Schwind et al., 2018) as well as human stimuli with the latter leveraging on theories of evolutionary aesthetics. These theories frame the UV effect as resulting from a mechanism for avoiding mates with low fitness as determined by the absence of physical markers of fertility, health, and youthfulness (MacDorman et al., 2009; MacDorman & Ishiguro, 2006).

Animacy and experience: Items measuring perceived animacy (Looser & Wheatley, 2010), responsiveness (Tinwell et al., 2014), and mind (Appel et al., 2016). These items relate to theories about how the perceived presence or absence of these qualities elicits the UV effect. For example, Gray and Wegner (2012) proposed that a machine having conscious experiences—or a human being lacking them—would be perceived as uncanny; the authors' creation techniques are broad: android robot videos, text about a supercomputer, and a photo of a man.

Anomaly: Items measuring an entity's perceived deviation from the norm. Anomaly items, such as *strange* or *weird*, are associated with *atypicality* theories. These theories predict that the UV effect is elicited by an entity whose features cause it to deviate strongly from its prototype (Kätsyri et al., 2015; Strait et al., 2017). Anomalies are easily created in images, where features can be moved, reflected, rotated, and scaled (e.g., Diel & MacDorman, 2021).

Disgust: Items measuring disgust, a predictor of the UV effect (Ho, MacDorman, & Pramono, 2008). These items relate to the theory that the UV effect results from an evolved mechanism for pathogen avoidance (MacDorman & Entezari, 2015).

Distinctive experience: Items measuring the UV effect as the subjective experience of *uncanniness* or *eeriness*, which may be correlated with fear, anxiety, and disgust (Bartneck et al., 2009a; Ho, MacDorman, & Pramono, 2008). This research conceives of the UV effect as an experience distinct from general psychological discomfort or anxiety. Gahrn-Andersen (2020) and Mangan (2015) have related the phenomenological study of the uncanny to the theories of Martin Heidegger and William James.

Familiarity: Items measuring the UV effect as feelings of unfamiliarity, based on Reichardt's (1978) translation of *shinwakan* as familiarity. Typically, in cognitive psychology, familiarity is contrasted with novelty: 0% familiarity is 100% novelty. However, when inspecting the *y*-axis of Mori's (2012) graph, the familiar–novel contrast leads to contradiction. On this interpretation, the bottom of the valley lies in negative familiarity, beyond 100% novelty, which cannot exist. One finds a different interpretation in Freud's (1919/2003) theory of the uncanny. To Freud, the uncanny is not the perception of something novel or unfamiliar. Rather, it is the recollection of something intimately familiar, perhaps from early childhood, that has long been estranged through repression (MacDorman & Entezari, 2015; MacDorman & Ishiguro, 2006). Freud asserts that repression transforms every emotional affect—including uncanniness—into anxiety (*Angst*).

General anxiety: Items measuring a state of anxiety or stress without relating it specifically to the subjective experience of the uncanny. The items are associated with theories based on category inhibition, cognitive conflict (Ferrey et al., 2015), and perceptual tension (Moore, 2012). Their use may reflect the assumption that the experiential quality of the UV effect is no more specific than the psychological discomfort caused by cognitive dissonance or cognitive load.

Interpersonal warmth: Items measuring the primary dimension of social perception, interpersonal warmth, which accounts for 53% of the variance in perceptions of social behaviors (Fiske, Cuddy, & Glick, 2007; Fiske, Cuddy, Glick, & Xu, 2002). This dimension is measured with positive affect items, like *likable*, *pleasant*, and *friendly*, which load on the same factor in factor analyses (Bartneck et al., 2009a; Ho & MacDorman, 2010). The construct is intended to measure how feelings about an entity change with its degree of human likeness. The dimension is roughly synonymous with affinity, the *y*-axis of Mori's (2012) graph, though as a construct *warmth* has been more thoroughly investigated. The use of *warmth* items to measure the UV effect is grounded in the assumption that *warmth* and *uncanniness* are inversely related. However, feelings of coldness—the low end of the scale—differ from feelings of uncanniness. For example, we might have warm feelings for the conductor (Tom Hanks) in *The Polar Express* (2004) while also having uncanny feelings because of the way he is computer animated. Furthermore, the generality of *warmth* items makes them susceptible to confounds. Stimulus evaluation could be influenced by, for example, background, clothing, color, narrative and framing, verbal and nonverbal behavior, interactivity, personality, relationships, and culture (Brink et al., 2019; Kennedy, 2014; Łupkowski, Rybka, Dziedzic, & Włodarczyk, 2018; MacDorman, 2019; Shin, Kim, & Biocca, 2019). Thus, *warmth* items do not indicate the UV effect but a related construct.

Threat: Items measuring a negative emotional response to dead animals, ranked by the species' similarity to living humans, motivated by theories that conceive of the UV effect as an evolved threat-avoidance mechanism (Moosa & Ud-Dean, 2010; Palomäki et al.,2018; Rosenthal et al., 2014). The entities could also appear threatening because of their ambiguity (McAndrew & Koehnke, 2016).

Trust: Numerical indicators of trust, such as the amount of money invested while playing a game, with a smaller investment indicating less trust. A decrease in trust could result from the UV effect in perceiving android robots or avatars. Mathur and Reichling (2016) relate this measure of trust to Hardin's (2002) theory of encapsulated interest: We trust those whose interest encapsulates our own. In their game, they raise the question of whether human players were really taking an intentional stance toward the robot or merely acting as if they were.

2 Methods

The lack of consensus in the UV literature, both theoretical and methodological, should now be evident. It motivates our meta-analysis, the first of its kind. We evaluate the effectiveness of stimulus creation techniques as well as affect and indirect measures. Based on the results, we propose empirically derived design principles for future research.

2.1 Inclusion criteria

The meta-analysis only included a study if it met the criteria below based on the information given:

Empirical study: The study contains the results of at least one data analysis conducted by its authors.

Representative participants: The study uses healthy adults, children, or infants. Excluded were studies restricted to a specific subgroup, such as people with autism spectrum disorder.

Relevant stimuli: The stimuli belong to at least one of the 10 creation techniques.

Adequate stimuli: The stimuli lack obvious confounds like noise created by editing images.

Affect or indirect measures: Affect measures include single-scale items or indices used to self-report an affective appraisal of the stimulus. Indirect measures include everything else. Studies with either or both were included.

Testing a UV hypothesis for statistical significance: The study has one or more hypotheses designed to test the UV effect. For each hypothesis, a test statistic is applied to the collected data. Studies with both significant and nonsignificant effects were included.

Appropriate variables: Testing for a change in an affect or indirect measure resulting from a change in human likeness or a related variable (e.g., realism, zoomorphism). Thus, all studies were experiments.

Effect size determinable: The study must give enough information to calculate an effect size and its variance.

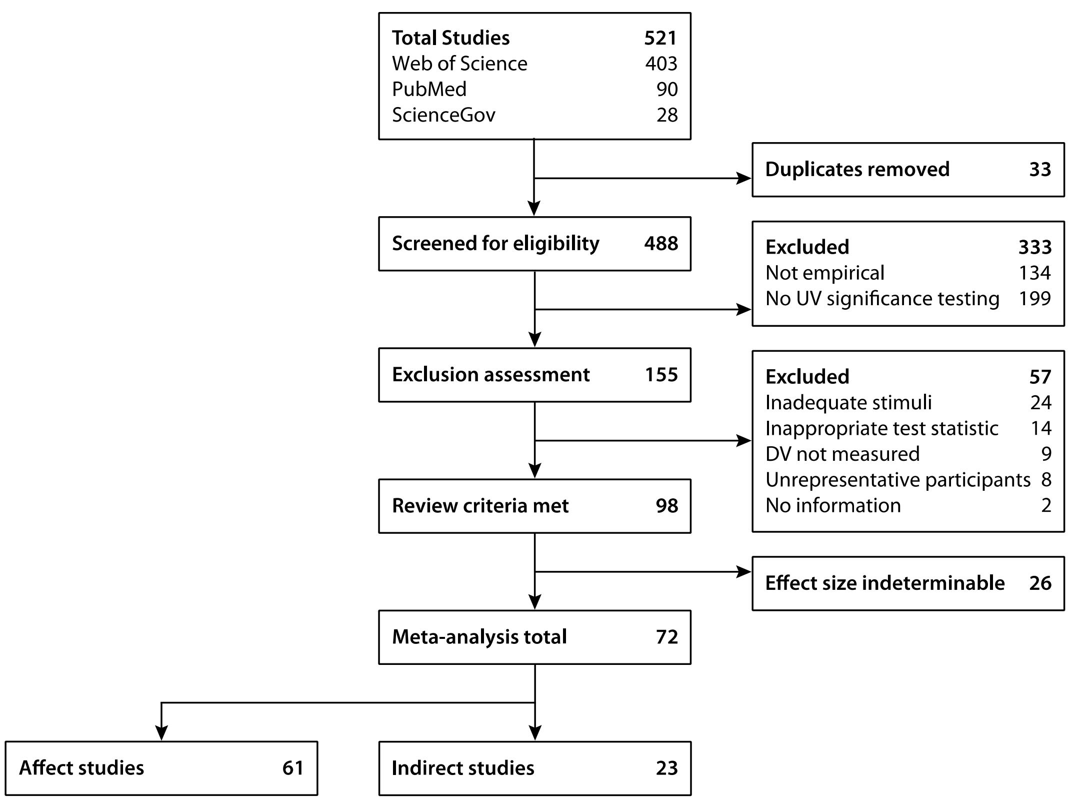

Figure 3. The flowchart depicts the process of study selection.

2.2 Study search and selection

In March 2021, we searched on PubMed, Science.Gov, and the Web of Science for papers with *uncanny valley* in their title, abstract, or keywords. After removing 33 duplicates, 488 studies remained of which 155 included UV significance testing (see Data Availability). Although 98 met other review criteria, only 72 had determinable effect sizes. These studies appeared in 56 papers published from 2008 to 2021. Figure 3 summarizes the article selection process.

From its description, we placed each IV operationalization under the best-fitting stimulus creation technique.

For DV operationalizations, single items were generally grouped separately. Nouns formed from adjectives were grouped with those adjectives (e.g., *eeriness* with *eerie*). The item *creepy* and semantic differential scales like *creepy–friendly* and *creepy–pleasant* were group as *creepy**. Affect measures were grouped separately from indirect measures. For example, the item *trustworthy* was counted as an affect measure, separate from trust behavior, an indirect measure. If a study used a negative variant of an often-used positive item, the item was grouped with the positive variant (e.g., *unpleasant* with *pleasant*). Indices used in multiple studies were counted as separate index items and marked with the suffix *-i* (e.g., those developed by Bartneck et al., 2009b; Ho and MacDorman, 2010, 2017; Schwind et al., 2018).

We then recorded or calculated effect sizes and effect size variances, labeling each with its corresponding IV and DV. If a study used more than one IV or DV operationalization, each effect size was recorded or calculated.

2.3 Data analysis

A random-effects model was selected for the meta-analysis because study populations and designs differed and affect and indirect measures were used in combination with different stimulus creation techniques. A three-level model was used with effect nested by study. The meta-regression for moderation analysis was performed using a mixed-effects model. The model was fitted by restricted maximum-likelihood estimation.

Effect size is reported here as Hedges' g. The effect size, its 95% confidence interval, and the number of measures from which it was derived, k, are all reported. Effect size is interpreted with small = 0.20, medium = 0.50, and large = 0.80 thresholds.

If three or more conditions were compared, such as robot, android, and human, two separate g's were calculated: one for the posited descent from the first peak in Mori's graph to the base of the valley and the second for the posited ascent from the base of the valley to the second peak. For convenience, the descent is denoted as the UV's *nonhuman* side and the ascent as the UV's *human* side.

The definition of an influential effect was adopted from Viechtbauer and Cheung (2010), as explained in the results section.

Moderator variables for the independent variable were the creation technique. Moderator variables for the dependent variable were (separately) the *side of the valley*, *side* × *valence* (positive or negative) × *measure type* (affect or indirect), *affect measure*, *indirect measure*, and *other construct*. Finally, *paper* was used as a moderator variable.

2.3.1 Effect size calculation

The meta-analysis used the standardized mean difference and its variance. Hedges' g was used to correct for the positive bias of Cohen's d in smaller studies,

$$g = d\left(1 - \frac{3}{4\,\mathrm{df} - 1}\right), \tag{1}$$

$$v_g = v_d\left(1 - \frac{3}{4\,\mathrm{df} - 1}\right)^2, \tag{2}$$

where df indicates the degrees of freedom (Borenstein et al., 2011). If a study did not report g, it was calculated from the means and standard deviations or by converting another reported measure of effect size. For within-group studies, which were the majority, d_{av} and v_{av} were used,

$$d_{\mathrm{av}} = \frac{m_1 - m_2}{\frac{1}{2}(s_1 + s_2)}, \tag{3}$$

$$v_{d_{\mathrm{av}}} = \frac{1}{n} + \frac{d^2}{2n}, \tag{4}$$

where n is the number of participants (Lakens, 2013). This approach leads to slightly wider confidence intervals than d for repeated measures. However, the calculation of d_{rm} requires the correlation between means, which no study reported. For ANOVAs, η^2 was first calculated:

$$\eta^2 = \frac{F \times \mathrm{df}_1}{F \times \mathrm{df}_1 + \mathrm{df}_2} \tag{5}$$

Next, to calculate g, η^2 was converted to d (Cohen, 1988):

$$d = 2\sqrt{\frac{\eta^2}{1 - \eta^2}} \tag{6}$$

R^2, Pearson's r, and Cramér's V were plugged into the same formula. For the t statistic, d was calculated for between-groups studies by imputing $r = 0.5$ in the formula

$$d = t\sqrt{\frac{2(1 - r)}{n}} \tag{7}$$

3 Results

The 72 studies in the meta-analysis employed 10 different stimulus creation techniques and 53 different measures, 39 of which were affect measures and 14 of which were indirect measures.

In total, 61 studies included affect measures, and 23 included indirect measures. The studies ranged in size from 10 to 1,311 participants with a median size of 64.5 and an interquartile range of 34 to 203.5. Of the 249 measured effects, 85 involve the nonhuman side of the UV, 71 involve the human side, and 93 involve both sides simultaneously.

The three-level meta-analysis model, including two outliers, revealed that the UV effect had a large effect size, $g = 0.95$ [0.76, 1.14], $p < .001$, $k = 249$, Akaike information criterion (AIC) =

724.92, $QE(248) = 10241.38$, $p < .001$, $QM(1) = 93.30$, $p < .001$. Excluding the two outliers, discussed below, increased the effect size, $g = 1.01$ [0.80, 1.22], $p < .001$, $k = 247$.

3.1 Three-level model

The meta-analysis often draws multiple effect sizes from the same paper and even from the same study. Thus, the effect sizes are not statistically independent (Cheung, 2019). To address this, we investigated different three-level models.

The model with the lowest estimated prediction error, excluding outliers, has *paper* as its higher-order grouping variable and *effect* as its nested lower-order grouping variable, $QE(246) = 9725.21$, $p < .001$, $QM(1) = 88.53$, $p < .001$. The model has lower estimated prediction error (paper/effect: $AIC = 675.17$) than the other three-level models (study/effect: $AIC = 683.05$, technique/effect: $AIC = 714.85$, measure/effect: $AIC = 715.20$). Its prediction error is significantly lower than two-level models (effect: $AIC = 717.57$, $p < .001$, paper: $AIC = 4915.67$, $p < .001$). Of the total variance, 38.53% is between-paper heterogeneity, 60.34% is within-paper heterogeneity (total $I^2 = 98.87$), and 1.13% is sampling error.

3.2 Bias

Figure 4(a) shows a funnel plot of effect sizes against their standard errors for meta-analysis. Since standard error is inversely proportional to sample size, larger studies appear at the top and smaller studies at the bottom. In the absence of bias, sampling error should distribute effect sizes randomly but symmetrically about their weighted mean. In the funnel plot, however, the effect sizes tend to increase with their standard errors. A regression test with standard error as the predictor variable and Hedges' g as the outcome variable indicated significant funnel plot asymmetry ($z = 6.72$, $p < .001$, $k = 249$).

Funnel plot asymmetry could result from publication bias because the meta-analysis relied on published data only. In general, studies reporting a significant effect are more likely to be published. If a true effect exists, a smaller study will require a larger effect size to reach significance. Moreover, given that large studies constitute a major commitment of resources, they are more likely to be published even if their effects are nonsignificant.

One approach to addressing bias is to limit the meta-analysis to larger studies and then to check whether bias is still present and whether the effect size is still large enough to be of substantive importance (Borenstein et al., 2009). We tried a version of this approach by excluding the effects with the largest standard errors and retesting for funnel plot asymmetry. After excluding 66 effects—that is, 27% of the total, as shown in Figure 4(b)—funnel plot asymmetry for the remaining effects became nonsignificant ($z = 1.95$, $p = .051$, $k = 183$). The effect size, however, was reduced 28%, $g = 0.68$ [0.51, 0.85], $k = 183$. Though smaller, it remains of substantive importance.

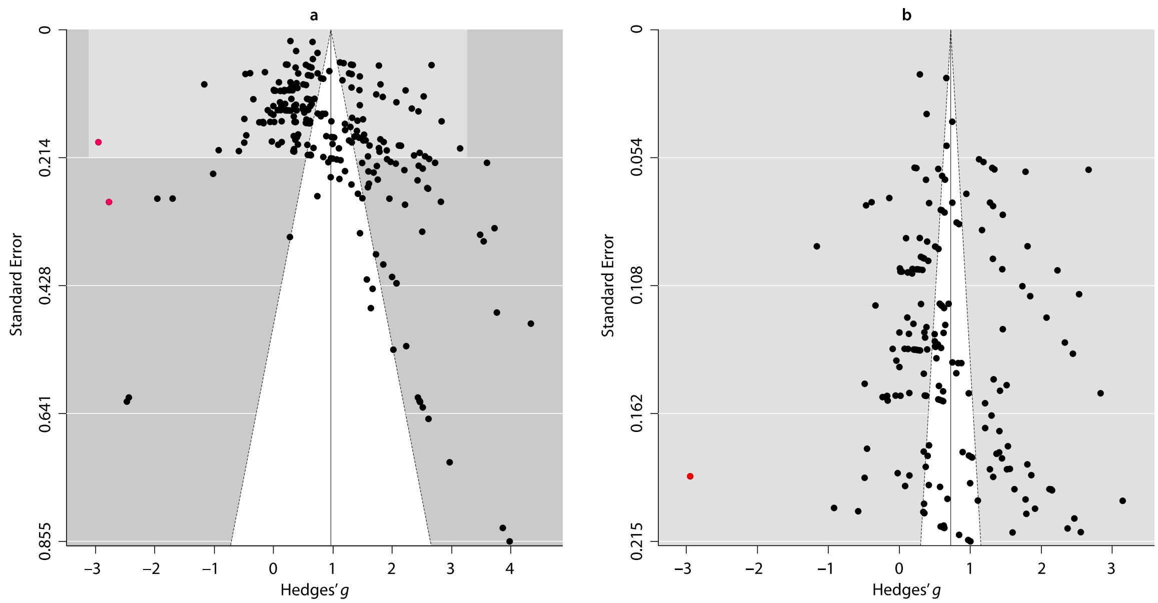

Figure 4. The funnel plot graphs effect sizes from the meta-analysis against their standard errors: (a) all standard errors; (b) the lowest 73% of standard errors. Influential effects are indicated in red.

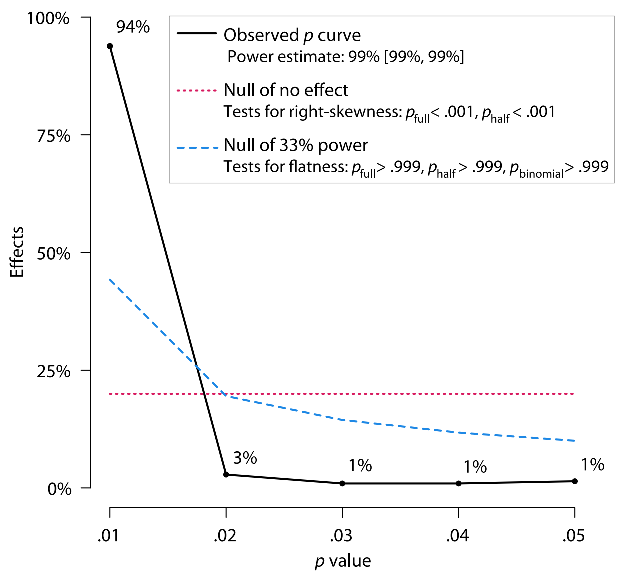

Figure 5. The p-curve for the meta-analysis's 249 effects.

Bias was next assessed by *p*-curve analysis. A plot of *p* values against percentage of effects should be flat if there is no effect and right skewed if there is one. A left skew indicates bias, a publication environment in which obtaining significance at the .05 level is incentivized, but lower *p* values are unnecessary. This could result from publication bias or from *p*-hacking, mining the data for patterns and then failing to control for multiplicity in reporting significance. Of 249 effects, $p \leq .05$ for 213 (86%), and $p \leq .025$ for 207 (83%). The right-skewness test, $p_{\text{binomial}} < .001$, $z_{\text{full}} = -73.80$, $p_{\text{full}} < .001$, $z_{\text{half}} = -72.50$, $p_{\text{half}} < .001$, was significant, which indicates a true effect (Figure 5). The flatness test was nonsignificant, $p_{\text{binomial}} > .999$, $z_{\text{full}} = 65.35$, $p_{\text{full}} > .999$, z_{half}

= 69.70, p_{half} > .999; thus, the test did not indicate insufficient power or the absence of a true effect. The power estimate is 0.99 [0.99, 1.00]. The tests were repeated, with similar results, for only the 66 effects with the largest standard errors. Thus, *p*-curve analysis supports the conclusion that the effect is true. It is not simply the result of publication bias or *p*-hacking.

3.3 Influential effects

Viechtbauer and Cheung (2010) proposed that an effect is influential if it meets one of the following four criteria:

$$|\text{DFFITS}| > 3\sqrt{\frac{p}{k-p}}, \tag{8}$$

where *p* is the number of model coefficients and *k* the number of effects, the Cook's distance,

$$D_i > \chi^2_{p,50\%}, \tag{9}$$

where *p* is the model's degrees of freedom, indicating the deletion if the *i*'th effect decreases the Mahalanobis distance between effects,

$$hat > \frac{3p}{k}, \text{ and any} \tag{10}$$

$$\text{DFBETA} > 1. \tag{11}$$

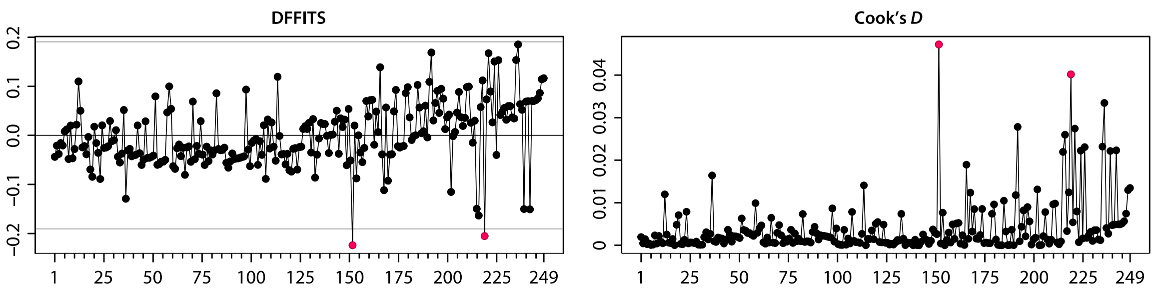

Figure 6. DFFITS and Cook's D for the effects in the meta-analysis, sorted from lowest to highest standard error. Influential effects are indicated in red.

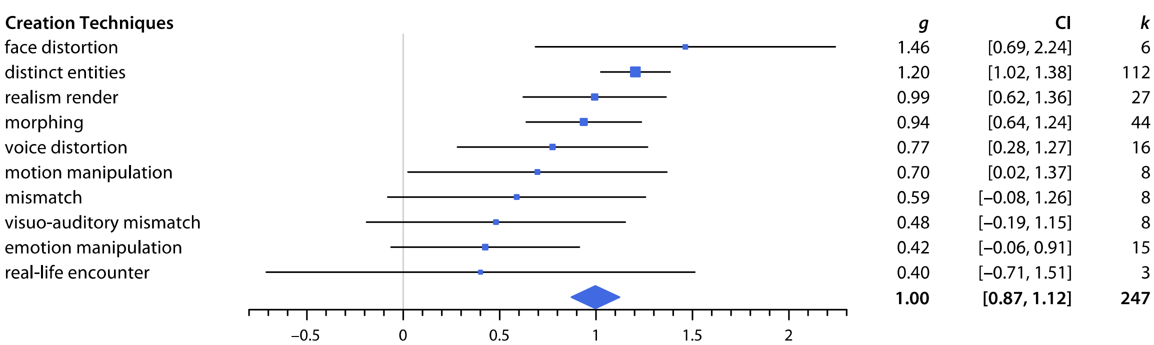

Figure 7. Creation technique is the moderator variable in the meta-regression model. For each of its values, Hedges' g, the 95% confidence interval, and number of effects (k) are listed. The position of the blue square depicts the effect size, and its relative size depicts the precision. The width of the diamond depicts the confidence interval of the summary effect size.

Two effects were identified as influential by the first two criteria (Figure 6), and both pertained to the UV's nonhuman side: Rosenthal et al.'s (2014) *unfamiliar-i*, $g = –2.95$, DFFITS = –0.224, D = 0.047, *hat* = 0.004, DFBETA = –0.224, and Wang et al.'s (2020) *alive*, $g = –2.77$, DFFITS = –0.205, D = 0.040, *hat* =0.004, DFBETA = –0.205. They were treated as outliers for reasons discussed below and included in analyses selectively.

3.4 Independent variable operationalizations

3.4.1 Moderator: Creation techniques

Moderation analysis was performed, excluding outliers, using a mixed-effects meta-regression model with effect as the random variable and creation technique as the moderator variable, AIC = 701.33, $QE(237) = 8984.08$, $p < .001$, $\tau^2 = 0.91$, $I^2 = 98.62$, $QM(10) = 272.53$, $p < .001$. *Face distortion* produced the largest effect size, followed by *distinct entities*, *realism render*, and *morphing* (Figure 7).

Distinct entities studies typically used stimuli that could have confounding effects (e.g., body language, facial expressions, lighting, viewing perspective). To reduce their risk, a few studies applied standards for stimulus selection—for example, full face shown in frontal or three-fourths aspect, resolution sufficient to generate a final image three inches in height at 100 dpi, and no other body parts visible (Brink, Gray, & Wellman, 2017; Mathur & Reichling, 2016). When only *distinct entities* studies with standardized stimuli were considered, three in total, g fell to 0.82 [–0.12, 1.77], $k = 4$, and the effect became nonsignificant, $p = .089$.

Four studies used nonhuman animal stimuli, $AIC = 32.95$, $QE(17) = 373.46$, $p < .001$, $QM(1) = 32.95$, $p < .001$ (MacDorman & Chattopadhyay, 2017; Schwind et al., 2018; Yamada et al., 2013). Their 18 effects were all significant, $g = 1.94$ [1.28, 2.60], $k = 18$. Stimulus operationalization techniques for animal stimuli were comparable with those for human stimuli, including *distinct entities* (Rativa et al., 2020; Takahashi et al., 2015), *emotion manipulation*, *face distortion*, *realism render* (Chattopadhyay & MacDorman, 2016; Schwind et al., 2018), and *morphing* (Yamada et al., 2013).

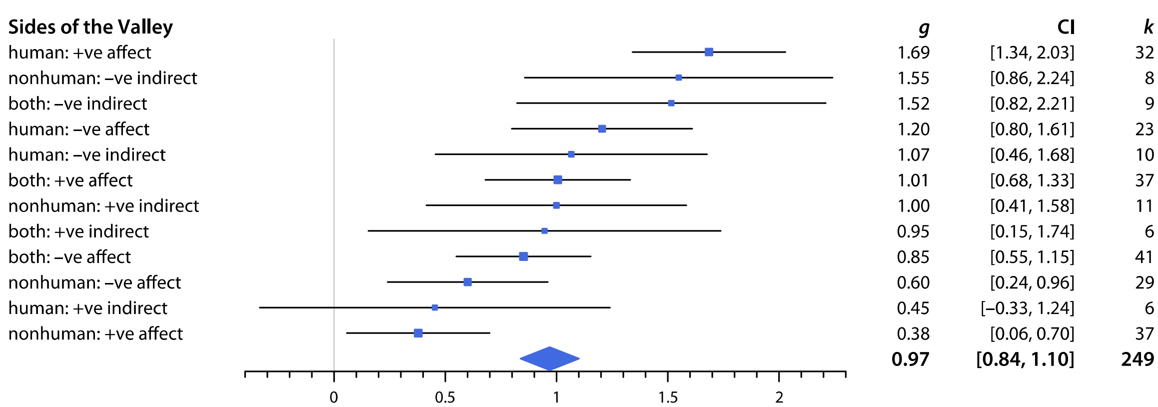

Figure 8. Side of the uncanny valley is the moderator variable in the meta-regression model.

3.5 Dependent variable operationalizations

3.5.1 Moderator: Side of the uncanny valley, valence, and type of measure

Moderation analysis was performed, including outliers, with *effect* as the random variable and *side of the valley* as the moderator variable, $AIC = 731.92$, $QE(246) = 9942.04$, $p < .001$, $\tau^2 = 1.00$, $I^2 = 98.80$, $QM(3) = 239.92$, $p < .001$. If possible, an effect size was calculated for each side of the uncanny valley. However, this was not possible for 37% of effect sizes, usually because the means and standard deviations were not reported. In these cases, a combined effect size for both sides of the valley was calculated (e.g., based on an F statistic). For the human side, $g = 1.34$ [1.10, 1.57], $p < .001$, and $k = 71$, for the nonhuman side, $g = 0.64$ [0.43, 0.86], $p < .001$, and $k = 85$, and for both sides, $g = 0.98$ [0.77, 1.19], $p < .001$, and $k = 93$. Thus, the effect size for the human side was more than double that of the nonhuman side.

To investigate this disparity, we repeated the analysis with *side* × *valence* (positive or negative) × *measure type* (affect or indirect) as the moderator variable (Figure 8). The combined value *human positive affect* had the largest affect size, $g = 1.69$ [1.34, 2.03], $p < .001$, $k = 32$ and *nonhuman positive affect* had the smallest. Thus, among all measures, positive affect measures were the most effective at measuring the human side of the valley and the least effective at measuring the nonhuman side. A Wald-type test revealed this difference in effectiveness was significant, $QM(12) = 276.73$, $p < .001$. For the human side, affect measures were more effective than indirect measures. For the nonhuman side, indirect measures were more effective than affect measures, and negative measures were more effective than positive ones.

3.5.2 Moderator: Affect measures

Moderation analysis was performed, excluding outliers, with *effect* as the random variable and *affect measure* as the moderator variable, $AIC = 537.05$, $QE(159) = 4544.64$, $p < .001$, $\tau^2 = 0.92$, $I^2 = 98.51$, $QM(38) = 247.70$, $p < .001$ (Figure 9). Indices producing effects that were larger than average include *threatening-i* (*threatening*, *eerie*, *uncanny*, *dominant*, *harmless*), *likable-i* (*pleasant*, *likable*, *attractive*, *familiar*, *natural*, *intelligent*), *aesthetics-i* (*ugly–beautiful*, *unaesthetic–aesthetic*), *familiarity-i* (*uncanny–familiar*, *freaky–numbing*), and *eeriness-i* (*dull–freaky*, *predictable–eerie*, *plain–weird*, *ordinary–supernatural*, *boring–shocking*, *uninspiring–spine-tingling*, *predictable–thrilling*, *bland–uncanny*, *unemotional–hair-raising*). Individual items include *reassuring*, *threatening*, *believable*, *appealing*, *acceptable*, *alive*, and *eerie*. However, when the two outliers are included, *alive* falls from the 12th highest effect size, $g = 1.19$ [0.33, 2.06], $p = .007$, $k = 5$, to the 29th, $g = 0.55$ [–0.27, 1.37], $p = .191$, $k = 6$, and is no longer significant. The other outlier, *unfamiliar-i* (*strange*, *unfamiliar*) appears last, $g = -2.95$ [–4.94, –0.95], $p = .004$, $k = 1$.

3.5.3 Indices and multiple scale analyses

A variety of terms have been used to measure different constructs underlying the UV effect. The relations among the terms can give insight into the UV effect's experiential quality. In studies with several terms, we investigated their intercorrelations to determine whether they reflect the UV effect or instead a related construct. Table A1 in the Appendix lists the interscale correlations observed in the reviewed research.

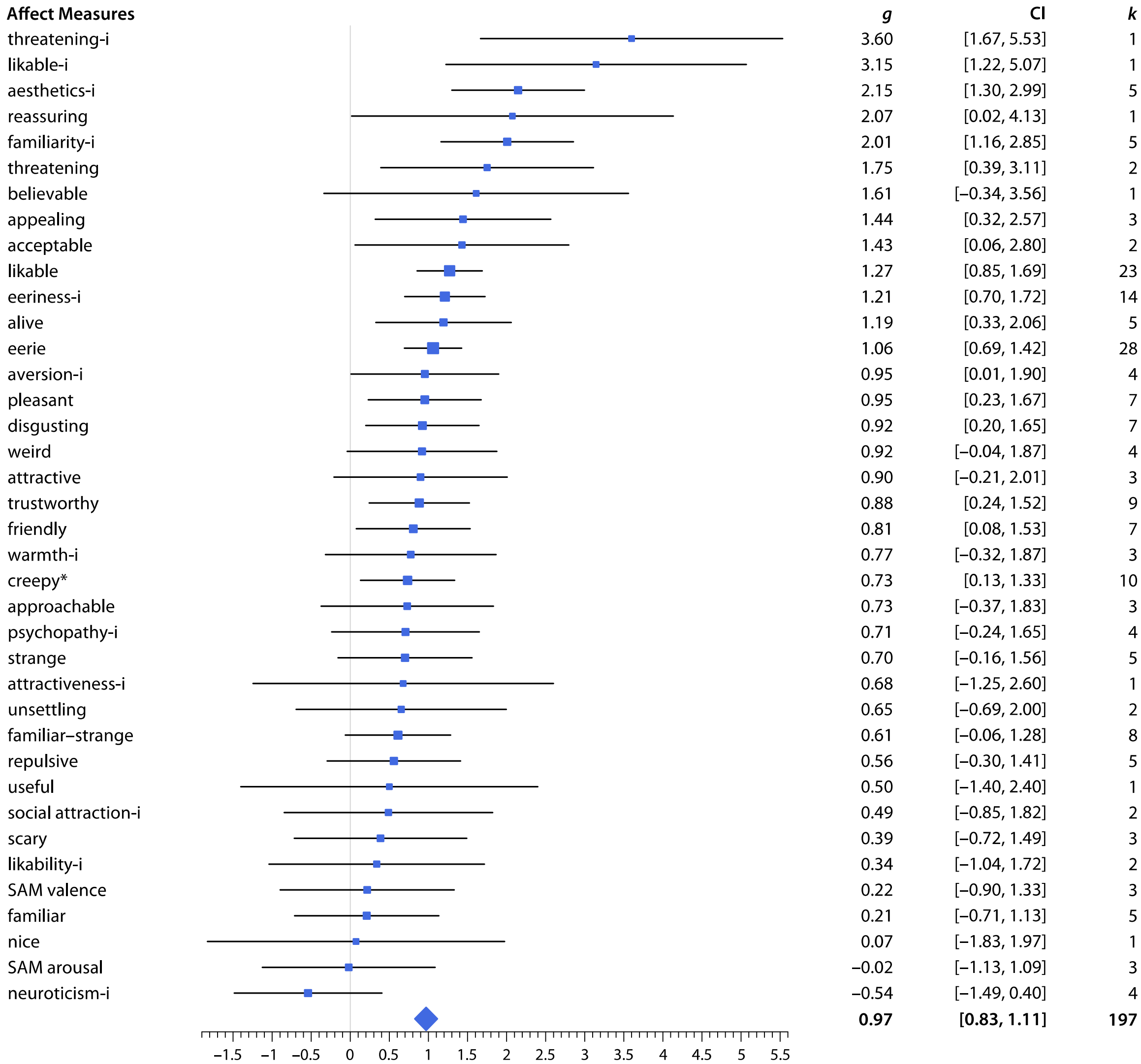

Figure 9. Affect measure is the moderator variable in the meta-regression model. Creepy* combines the item creepy with scales including the term, such as creepy–pleasant and creepy–friendly.

As a measure of reliability, 15 studies in the meta-analysis reported the Cronbach's α of the indices used. Ho and MacDorman's (2010, 2017) *eeriness* and *warmth* indices and their derivations were generally reliable. Distinctive experience terms (e.g., *creepy*, *eerie*, and *uncanny*) tended to load on the same factor (e.g., Destephe et al., 2015; Lischetzke et al. 2017). In a principal component analysis (PCA), the items *uncanny* and *eerie* loaded on the same component as threat-related items, and the items *strange* and *unfamiliar* as anxiety-related items (Rosenthal–von der Pütten & Krämer, 2014; Ho, MacDorman, & Pramono, 2008, found fear and disgust to be stronger predictors of *eerie* and *creepy* than anxiety). In a similar vein, removing strange from an index consisting of *eerie*, *unsettling*, and *strange* improved its reliability (Kätsyri, Mäkäräinen, & Takala, 2017). This indicates *uncanniness* and *strangeness* may be different constructs.

Finally, *likable*, *friendly*, *pleasant*, and other *warmth* items typically comprise reliable indices (e.g., Kätsyri, Mäkäräinen, & Takala, 2017; Rosenthal–von der Pütten & Krämer, 2014; Tung, 2016), which indicates an *interpersonal warmth* construct for the tested stimuli (e.g., Bartneck et al., 2009a).

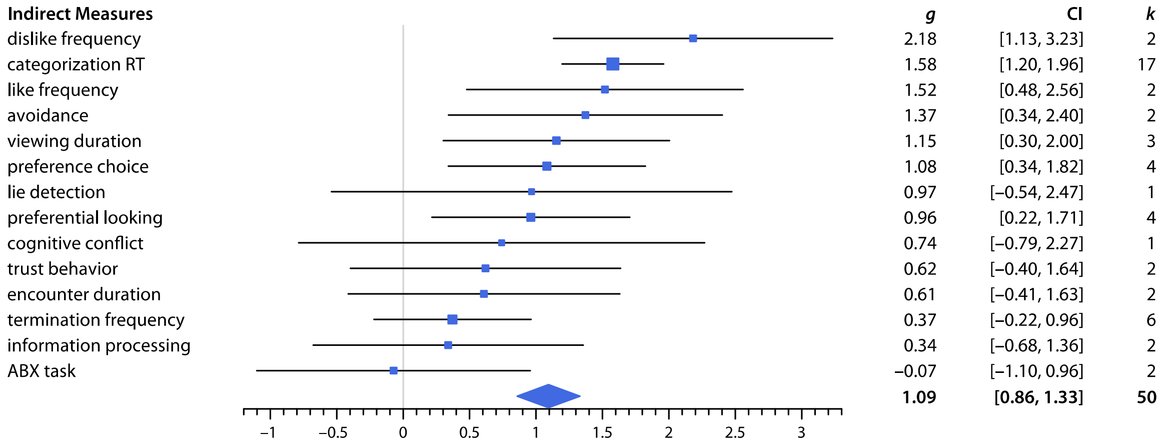

Figure 10. Indirect measure is the moderator variable in the meta-regression model.

3.5.4 Moderator: Indirect measures

Moderation analysis was performed, excluding outliers, with *effect* as the random variable and *indirect measure* as the moderator variable (Figure 10). *Dislike frequency*, which indicates the number of times disliked, had the largest effect size (Strait et al., 2019), followed by *categorization reaction time* (Carr et al., 2017; Cheetham & Jäncke, 2013; MacDorman & Chattopadhyay, 2017; Wang & Rochat, 2017; Yamada et al., 2013), *like frequency* (Strait et al., 2019), *avoidance* behavior attributions to uncanniness (Perez et al., 2020), *viewing duration* (Strait et al., 2015, 2019), *preference choice* in a two-alternative forced-choice categorization task (Feng et al., 2018; Prakash & Rogers, 2015), and *preferential looking*, that is, preferring to view one stimulus more than another (Matsuda et al., 2015; Nitta & Hashiya, 2021).

Nonsignificant effect sizes include *lie detection*, that is, frequency of rating a statement as a lie (McDonnell & Breidt, 2010), *cognitive conflict*, operationalized as number of reversals of direction when moving a stimulus with a mouse pointer towards one of two categories (Weis & Wiese, 2017), *trust behavior*, specifically the amount of money entrusted with an entity in an investment game (Mathur & Reichling, 2016), *encounter duration*, that is, viewing duration until the participant terminates the encounter (Perez et al., 2020), *termination frequency*, measured by the number of times terminated (Perez et al., 2020; Strait et al., 2015, 2017, 2019), *information processing* about an entity, as indicated by the number of personality judgments made (Shin, Kim, & Biocca, 2019), and *ABX task*, which entails visual same–different discriminations (Cheetham et al., 2014).

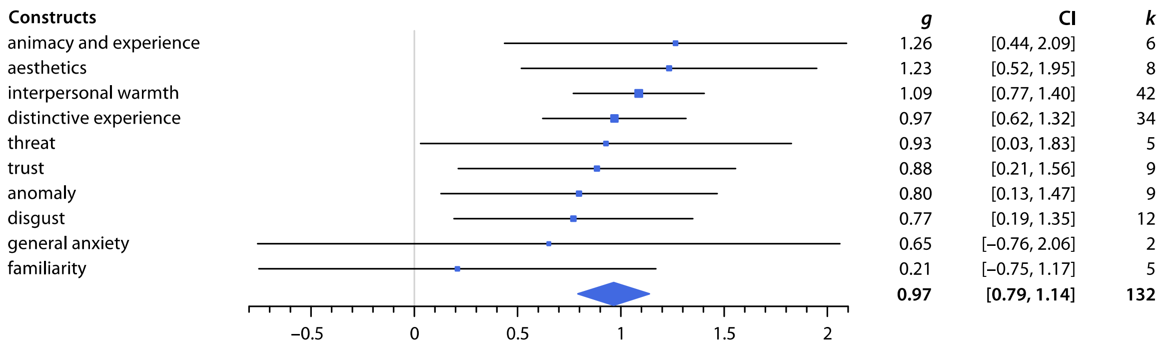

Figure 11. Construct is the moderator variable in the meta-regression model.

3.6 Other constructs

After grouping measures by other UV construct, moderation analysis was performed, excluding outliers, with *effect* as the random variable and *other construct* as the moderator variable, AIC = 386.28, $QE(122) = 2999.63$, $p < .001$, $\tau^2 = 1.02$, $I^2 = 98.29$, $QM(10) = 119.67$, $p < .001$ (Figure 11). *Animacy and experience* had the largest effect size, $g = 1.26$ [0.44, 2.09], $p = .003$, $k = 6$. However, if outliers are included, this construct falls from first to eighth and becomes nonsignificant, $g = 0.70$ [–0.10, 1.51], $p = .088$, $k = 7$. Other constructs with significant effects, in decreasing order of effect size, were *aesthetics*, *interpersonal warmth*, *distinctive experience*, *threat*, *trust*, *anomaly*, and *disgust*. *General anxiety* and *familiarity* had nonsignificant effects.

3.7 Papers

For reference, a moderation analysis was performed, excluding outliers, with *effect* as the random variable and *paper* as the moderator variable, AIC = 585.95, $QE(191) = 5058.35$, $p < .001$, $\tau^2 = 0.61$, $I^2 = 98.05$, $QM(56) = 552.95$, $p < .001$ (Figure 12).

3.8 Data availability

The meta-analysis was performed in the *R* statistical computing environment with the metafor package. The *p*-curve analysis and variance distribution analysis of the three-level model were performed with the dmetar package. The remaining *R* packages were devtools, forestplot, ggplot2, and readxl. The dataset, R script, and other supplementary materials are available at https://doi.org/10.17605/osf.io/57sme.

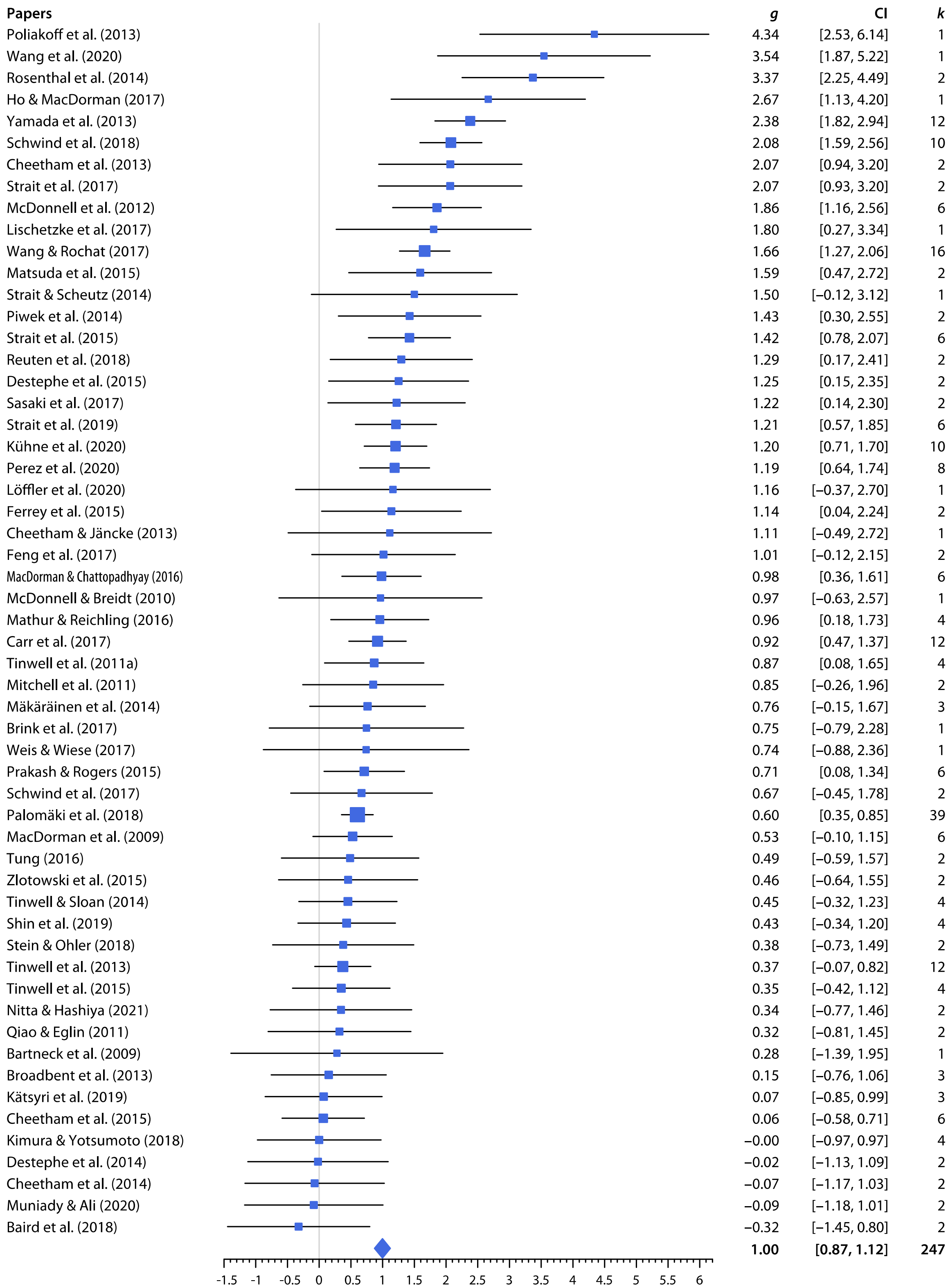

Figure 12. Paper is the moderator variable in the meta-regression model.

4 Discussion

4.1 Independent variable operationalizations

Among all the stimulus creation techniques, *face distortion* produced the largest effect size, followed by *distinct entities*, *realism render*, *morphing*, *voice distortion*, and *motion manipulation*. Techniques producing a nonsignificant effect include *mismatch*, *visuo-auditory mismatch*, *emotion manipulation*, and *real-life encounter*, though *real-life encounter* was based on only one paper. Nonhuman animal stimuli performed well. Our evaluation of stimulus creation techniques is summarized in Table A2 of the Appendix.

Face distortion was only tested in four of the papers reviewed (Feng et al., 2018; MacDorman et al., 2009; Mäkäräinen et al., 2014; Schwind et al., 2018). Nevertheless, it is a promising technique to explore configural processing theories (Diel & MacDorman, 2021).

Distinct entities were used in 46% of significance tests (114 out of 249), more than any other technique. This creation technique has greater ecological validity than all techniques except—at least for robots—*real-life encounter*. However, stimuli in these studies typically varied in body language, facial expression, familiarity, gaze direction, lighting, perspective, and other aspects. These potential confounding variables indicate a lack of experimental control, which could limit the generalizability of the results (Kätsyri, Förger, Mäkäräinen, & Takala, 2015; Kätsyri, de Gelder, & Takala, 2019). This interpretation aligns with our results. When the moderation analysis was limited to studies using standardized stimuli, *distinct entities* produced a nonsignificant effect.

Although *morphing* produced a large effect size in the meta-analysis, it was nonsignificant for 8 out of 44 effects. Nonsignificance may stem from the choice of endpoint stimuli. Studies that did not find a UV effect used endpoint stimuli with the same shape, such as a human face and a matching avatar face (Cheetham et al., 2015; Kätsyri, de Gelder, & Takala, 2019; the same issue arises for *realism render*, MacDorman & Chattopadhyay, 2016). By contrast, studies that did find a UV effect used morphologically different endpoint stimuli to produce a robot-to-human, animal-to-human, or cartoon-to-real transition (Ferrey et al., 2015; Lischetzke et al., 2017; Palomäki et al., 2018; Sasaki, Ihaya, & Yamada, 2017).

Creating stimuli from insufficiently distinct endpoint images may result in a morphing sequence with too narrow a range in human likeness to include the uncanny valley part of the graph. For example, although animals and robots have facial proportions that are atypical for humans, they are not judged by human standards. Morphing them with human faces may elicit human-specific processing, heightening sensitivity to those features that still deviate from human proportions, thus eliciting the UV effect. This effect could not occur if the facial proportions of the low human likeness endpoint stimuli were already human (e.g., human avatars). Thus, it is possible that, for morphing stimuli to elicit a UV effect reliably, they must distort an entity's configural pattern, which would support theories predicting the UV effect results from configural processing (Chattopadhyay & MacDorman, 2016; Diel & MacDorman, 2021; Kätsyri, 2018).

Alternatively, the large effect sizes for endpoint stimuli that differ greatly in their morphology may be an unintended consequence of the creation technique. Endpoint stimuli like robots and dolls tend to be attractive because they are the product of design. Human beings, though not designed, tend to find each other attractive because their faces and bodies co-evolved with their perceptual systems. In this context, attractiveness serves a purpose: It supports mate bonds and parental bonds (see Kozak, Head, Lackey, & Boughman, 2013; Wyman, Charlton, Locatelli, & Reby, 2011). However, intermediate stimuli in a morphing sequence neither evolved nor were designed to be perceived as anything. This arbitrariness could heighten their uncanniness.

We advise researchers to avoid using similar endpoint images when creating stimuli through morphing, or to use such techniques as morphing different regions of the face in different morphing steps (Seyama & Nagayama, 2007). However, it is also important to avoid creating strange or ghostly artifacts that could appear eerie for reasons other than their being intermediate in human likeness (discussed in MacDorman & Chattopadhyay, 2016). The effect of endpoint stimulus choice on the UV effect is a topic for investigation.

In their review, Wang, Lilienfeld, and Rochat (2015) found evidence against the UV effect comes from studies using *distinct entities*, while evidence for the UV effect comes from studies using *morphing*. The reason is perhaps that Wang and colleagues cited studies our analysis excluded for not using a test statistic (Hanson et al., 2005) or for having image noise (e.g., one face with two sets of hair, Seyama & Nagayama, 2007). In addition, several *distinct entities* studies with supportive results were published after their review (Brink, Gray, & Wellman, 2017; Jung & Cho, 2018; Kätsyri, de Gelder, & Takala, 2019; Mathur & Reichling, 2016; Mathur et al., 2020; Palomäki et al., 2018; Strait et al., 2017).

Finally, Wang, Lilienfeld, and Rochat (2015) criticizes using *face distortion* as an independent variable because face distortion differs from human likeness. However, our review found *face distortion* can elicit UV-specific subjective experiences (e.g., Mäkäräinen et al., 2014). Moreover, our meta-analysis found a significant UV effect in perceiving animal stimuli (e.g., Löffler et al., 2020; Schwind et al., 2017, 2018). Thus, human likeness alone cannot predict the range of observed UV effects. A more encompassing DV conceptualization, like norm deviation, would predict a broader range of UV effects. However, norm deviation is not necessarily uncanny. Sometimes it does harm aesthetics but rather improves it (e.g., supernormal stimuli, Diel & MacDorman, 2021).

4.2 Dependent variable operationalizations

The effect size of the uncanny valley's human side was more than double that of its nonhuman side. This difference may seem to reflect Mori's graph because the second peak is higher than the first. However, we also noted that, among all measures, positive affect produced the largest effect sizes for the human side and the smallest for the nonhuman side. Thus, a more plausible explanation is that positive affect is a poor measure of the UV effect.

Setting aside the miraculous and the extraterrestrial, people tend to perceive human beings as superior to nonhuman entities. This applies to stimuli appearing in UV experiments to date, such as robots, animals, and dolls. Perceived limitations in present-day human artifacts or other species reinforce our ingroup bias, rooted in our common identity, to privilege the human (MacDorman & Entezari, 2015; Mitchell et al., 2011a). Humans are often seen as more appealing, attractive, friendly, likable, pleasant, reassuring, and warm than nonhuman alternatives, not to mention more cultured, intelligent, and sociable. We can immediately see why positive affect measures are poor for measuring the UV effect because, despite how uncanny an android may appear, it will still appear more lifelike and less unfamiliar than a mechanical-looking robot of a novel design. Thus, it is important to focus on effective measures for the uncanny valley's nonhuman side: negative affect measures and positive indirect measures.

The effectiveness of negative affect measures like *eerie*, *creepy*, *threatening*, and *disgusting* align with the view that the UV effect is characterized by a distinctive experience of uncanniness rather than an overall decrease in positive affect (e.g., Ho, MacDorman, & Pramono, 2008; Mangan, 2015; Redstone, 2013). This negative experience may still reduce positive affect, though indirectly (Patrick & Lavoro, 1997).

The most frequently used item was *eerie* (e.g., Ho & MacDorman, 2010, 2017; Kätsyri, de Gelder, & Takala, 2019). Other negative items included *creepy*, *disgusting*, *repulsive*, *strange*, *threatening*, and *weird*. Concordantly, positive items with the largest effect sizes were nonspecific, such as *interpersonal warmth* items (*likable*, *pleasant*) or *familiar* (e.g., MacDorman & Ishiguro, 2006). Despite a correlation between the UV effect and feelings of disgust (e.g., Ho, MacDorman, & Pramono, 2008; MacDorman & Entezari, 2015), the item *repulsive* was nonsignificant.

Among indirect measures, *dislike frequency* produced the largest effect size, followed by *categorization RT*, *like frequency*, *avoidance*, *viewing duration*, *preference choice*, and *preferential looking*. Indirect measures, such as performance measures, are not without their limitations. Although some research uses performance measures to quantify a construct related to, but distinct from, the UV effect, other research claims they measure the UV effect itself (e.g., Lewkowicz & Ghazanfar, 2012; Matsuda et al., 2015). Measures like *preferential looking* and *preference choice* reflect general avoidance behavior, which could be elicited by the UV effect or by extraneous factors that must be controlled for, such as an ugly appearance or inhospitable disposition. Furthermore, most studies measuring performance omitted affect. Those that measured it tended to find a UV effect for affect but not for performance (Strait et al., 2015; Strait, Urry, & Muentener, 2019; for the opposite case, see Wang & Rochat, 2017).

These findings point to broader issues with measurement in UV research: First, many studies do not measure affect, but they should endeavor to do so insofar as it is possible. It is better to avoid relying solely on task performance measures (e.g., categorization RT, Cheetham, Suter, & Jäncke, 2011; Cheetham et al., 2013; Cheetham, Suter, & Jäncke, 2014; Chen, Russell, & Nakayama, 2010; Saygin, Chaminade, Ishiguro, Driver, & Frith, 2012; avoidance or preference, Lewkowicz & Ghazanfar, 2012; Matsuda et al., 2012; Steckenfinger & Ghazanfar, 2009). The reason is that we cannot infer affect and its influence on motivation solely from nonaffective behavior, though we can code it from displays of emotion. For example, in a study that used *termination frequency* to measure the UV effect, "the stimulus was boring" had a larger effect size than "the stimulus was unnerving" (Strait et al., 2015; Strait, Urry, & Muentener, 2019). However, *boring* has never been considered the dependent variable in Mori's graph. In addition, task performance measures can diverge from affect measures (MacDorman & Chattopadhyay, 2016, 2017; Mathur et al., 2020). Research should aim to validate performance measures by testing their specificity for the UV effect.

Second, although *likability*, *pleasantness*, and other nonspecific items used to measure overall affect tend to correlate with UV-specific items, they do not capture the experiential quality of the UV effect. Thus, unrelated factors could cause them to increase or decrease. This makes nonspecific items more susceptible to confounding variables. Perceptual variables that can influence stimulus evaluation include attractiveness (Ho & MacDorman, 2010, 2017; Principe & Langlois, 2011), atypical (Kätsyri et al., 2015; Strait et al., 2017), disgusting (Curtis, de Barra, & Aunger, 2011), or misaligned features (MacDorman & Chattopadhyay, 2016), background (Łupkowski et al., 2019), color (Kennedy, 2014; Valdez & Mehrabian, 1994), morphing artifacts (MacDorman & Chattopadhyay, 2016), realism (McDonnell et al., 2012), and size (Cesarei & Codispoti, 2006). These variables tend to be automatic and stimulus-driven. Perceptual-cognitive variables include categorization difficulty (Cheetham et al., 2013; Yamada et al., 2013), expectation violation (Saygin et al., 2012), frequency (Burleigh & Schoenherr, 2015; Moreland & Zajonc, 1982), inhibitory devaluation (Ferrey, Burleigh, & Fenske, 2015; Weis & Wiese, 2017), and multimodal mismatch (Mitchell et al., 2011b; Tinwell et al., 2015). Social variables include animacy (Koldewyn, Hanus, & Balas, 2014; Mäkäräinen et al., 2014), context (Jung & Cho, 2018), facial expressions (Paulus & Wentura, 2015; Tinwell et al., 2011), mind (Gray & Wegner, 2012), narrative structure (MacDorman, 2019), outgroup membership (Hugenberg, 2005), and perceived warmth or competence (MacDorman, 2019). Thus, studies should include UV-specific measures to mitigate potential confounds.

Third, even when UV-specific measures are used, they can be influenced by the flow of the interaction and its narrative structure (Dai & MacDorman, 2018). Thus, it may be necessary to test for the UV effect before the interaction begins.

Fourth, the UV effect is correlated with fear, anxiety, and disgust (Ho, MacDorman, & Pramono, 2008). Thus, a UV measure should be able to discriminate UV stimuli from non-UV stimuli that elicit similar emotions. However, discriminant validity has not yet been demonstrated for a UV measure.

Fifth, regardless of the strength of a change in affect, at least three stimulus conditions are necessary to produce measurements that could fit a *U*-shaped curve—the valley part of Mori's graph. Even if those measurements fit, a dip in a measure like *interpersonal warmth* could occur for a myriad of reasons other than the UV effect. Thus, experimental control is vital.

Sixth, what *eeriness* is and which situations elicit it has not been specified precisely. Redstone (2013) proposed that *eeriness* is elicited when the ontological nature of a stimulus is unclear. Langer and König (2018) differentiate between *eeriness* (which they assert is a fear-related response to humanoid entities) and *creepiness* (an anxiety-related response to novel or unpredictable people or situations). However, these claims are untested. In general, UV research lacks a common definition and conceptualization of the UV effect.

4.3 Limitations

4.3.1 Study exclusion

This meta-analysis excluded a wide range of impactful UV studies that were never intended to replicate a UV curve. For example, Gray and Wegner (2012) found the UV effect was elicited by a conscious machine or the philosophers' zombie (a person lacking conscious experience). Their findings were replicated by Appel and colleagues (2020). Schein and Gray (2015) found that, among facial features, the UV effect was especially sensitive to the manipulation of the eyes. The review also excluded specific subgroups and nonhuman primates. For example, Steckenfinger and Ghazanfar (2009) found a UV effect in macaque monkeys. The meta-analysis also excluded studies on the neurophysiological correlates of the perception of humanlike appearance or behavior, which shed light on the neural mechanisms underlying the UV effect (e.g., Saygin et al., 2011; Urgen et al., 2018).

The meta-analysis excluded interaction effects for simplicity. However, these effects have elucidated the UV effect. For example, Green and colleagues (2008) found an interaction between the degree of *face distortion* and *realism render* by showing that sensitivity to acceptable facial proportions increased as the stimulus appeared more human. Similarly, Mäkäräinen and colleagues (2014) showed that the strangeness of faces with exaggerated expressions increased as faces were rendered more realistically. Both studies indicate realism increases the perceiver's sensitivity to human features. Thus, deviations from norms are more likely to be noticed and perceived as uncanny in realistic representations. Sensitivity increases with realism logistically (*S*-shaped curve), not linearly, indicating a perceptual magnet effect (Chattopadhyay & MacDorman, 2016) like the one found for animacy (Looser and Wheatley, 2010). In a similar vein, Deska and colleagues (2017) found that the perception of a mind occurs when a face appears nearly human and is processed configurally (cf. Gray & Wegner, 2012; Tinwell et al., 2013).

Smaller studies, which require a larger effect size to obtain significance, tended to have larger effect sizes in our meta-analysis. Specifically, the average effect size of smaller studies, those in the quartile with the largest standard errors, was more than double that of the other three quartiles. Typically, inflated effect sizes in smaller studies are explained by publication bias or *p*-hacking. Publication bias results from unpublished or unreported nonsignificant effects in a meta-analysis, and *p*-hacking is the failure to control for multiplicity in significance testing. However, *p*-curve analysis found no signs of publication bias or *p*-hacking.

Twenty-six of 98 studies that met selection criteria, including significance testing, were excluded from the meta-analysis because they provided insufficient information to calculate effect sizes. This issue arose mainly for nonsignificant effect sizes. Nevertheless, the field has shown interest in nonsignificant and contrary effects, and papers reporting them have been well-cited (e.g., Cheetham, Suter, & Jäncke, 2014; Thompson, Trafton, & McKnight, 2011). Because this paper focuses on comparing methodologies, bias affecting relative comparisons between effect sizes is more worrisome than bias affecting their absolute magnitude.

4.3.2 Diverse methodologies

The diversity of UV methodologies impeded the meta-analysis. The volume of IV–DV combinations complicated the interpretation of effect sizes for creation techniques and for measures, especially for IV–DV combinations used in only a few studies. Precision in meta-regression requires having enough combinations in each cell. At least five is one rule of thumb (Borenstein et al., 2009). However, three of 10 techniques, 23 of 39 affect measures, and 12 of 14 indirect measures were used in fewer than five studies. The variety of experimental designs and other study-specific variables also complicates interpretation of the results. To draw conclusions about techniques and methods simultaneously requires enough significance tests or effect sizes to make comparisons (Lay, Brace, Pike, & Pollick, 2016). Future research could give priority to the validation of rarely used methods.

5 Conclusion

This is the first meta-analysis on the UV effect. We used meta-regression to evaluate the methods used to operationalize the axes of Mori's graph. Our findings provide a methodological foundation for UV research. After discussing the conceptual foundations of the uncanny valley, we have presented successful research methodologies and raised methodological concerns.

5.1 Recommendations

We end by proposing the following design principles for stimulus creation techniques and measures in UV research:

Items that measure the UV experience as a distinct experience of uncanniness, such as *uncanny* and *eerie*, or of strangeness, such as *weird* or *strange*, are preferred to nonspecific items. They also have face validity. In this vein, negative items are preferred to positive ones. Negative items can always be reverse scaled to plot the valley.

Affect or preference measures are necessary to assess the UV effect. Although indirect measures may complement them, a study should not rely solely on indirect measures, if possible. The validity of performance measures warrants further investigation.

The stimulus creation techniques producing the largest effect sizes were *face distortion*, *distinct entities*, *realism render*, and *morphing*.

A drawback of *morphing* is that, if the endpoint images are too similar, the x-axis may not include the uncanny valley. *Morphing* that disrupts the configural pattern may produce a larger effect; however, it should avoid creating visual artifacts from the morphing process. How best to approach *morphing* is a topic for future research.

Useful stimulus creation techniques include distorting facial features, rendering at different realism levels, and using different emotional expressions. Their choice depends on theoretical considerations and the research question. Further investigation is needed on *realism rendering* and how it influences UV-specific negative measures compared with nonspecific positive measures.

When using *distinct entities*, researchers should apply standards for stimulus selection (e.g., similar size, perspective, facial expression, and lighting). The effect of stimulus standardization on the UV effect also warrants investigation.

REFERENCES

Markus Appel, David Izydorczyk, Silvana Weber, Martina Mara, and Tanja Lischetzke. 2020. The uncanny of mind in a machine: Humanoid robots as tools, agents, and experiencers. *Computers in Human Behavior, 102* (Jan. 2020), 274–286. https://doi.org/10.1016/j.chb.2019.07.031

Markus Appel, Silvana Weber, Stefan Krause, and Martina Mara. 2016. On the eeriness of service robots with emotional capabilities. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. IEEE Press, 411–412.

Alice Baird, Emilia Parada-Cabaleiro, Simone Hantke, Felix Burkhardt, Nicholas Cummins, and Björn Schuller. 2018. The perception and analysis of the likeability and human likeness of synthesized speech. *Proceedings of Interspeech 2018* (Sep. 2018), 2863–2867. https://doi.org/10.21437/Interspeech.2018-1093

Christoph Bartneck, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2009a. My robotic Doppelgänger: A critical look at the uncanny valley theory. *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication* (Nov. 2009) (RO-MAN, pp. 269–276). Toyama, Japan. https://doi.org/10.1109/roman.2009.5326351

Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009b. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics* 1, 71–81. https://doi.org/10.1007/s12369-008-0001-3

Kimberley A. Brink, Kurt Gray, and Henry M. Wellman. 2017. Creepiness creeps in: Uncanny valley feelings are acquired in childhood. *Child Development*, 90, 4 (Jul. 2019), 1202–1214. https://doi.org/10.1111/cdev.12999

Michael Borenstein, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein. 2009. *Introduction to meta-analysis*. Hoboken, NJ: Wiley.

Elizabeth Broadbent, Vinayak Kumar, Xingyan Li, John Sollers, III, Rebecca Q. Stafford, and Bruce A. MacDonald. 2013. Robots with display screens: A robot with a more humanlike face display is perceived to have more mind and a better personality. *PLOS One*, 8, 8 (Aug. 2013), 1–10. https://doi.org/10.1371/journal.pone.0072589

Tyler J. Burleigh and Jordan R. Schoenherr. 2015. A reappraisal of the uncanny valley: Categorical perception or frequency-based sensitization? *Frontiers in Psychology*, 5 (Jan. 2015), 1488. https://doi.org/10.3389/fpsyg.2014.01488

Tyler J. Burleigh, Jordan R. Schoenherr, & Guy L. Lacroix. 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 (May 2013), 759–771. https://doi.org/10.1016/j.chb.2012.11.021

Colleen Carpinella, Alisa Wyman, Michael Perez, and Steven Stroessner. 2017. The Robotic Social Attributes Scale (RoSAS): Development and validation. *ACM/IEEE International Conference on Human–Robot Interaction.* (pp. 254–262). New York, NY, USA. https://doi.org/10.1145/2909824.3020208

Evan W. Carr, Galit Hofree, Kayla Sheldon, Ayse P. Saygin, and Piotr Winkielman. 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 (Jan. 2017), 651–666. https://doi.org/10.1037/xhp0000304

Andrea de Cesarei and Maurizio Codispoti. 2006. When does size not matter? Effects of stimulus size on affective modulation. *Psychophysiology*, 43, 2 (Mar. 2006), 207–215. https://doi.org/10.1111/j.1469-8986.2006.00392.x

Thierry Chaminade, Jessica K. Hodgins, and Mitsuo Kawato. 2007. Anthropomorphism influences perception of computer-animated characters' actions (Sep. 2007). *Social Cognitive and Affective Neuroscience*, 2, 3, 206–216. https://doi.org/10.1111/j.1469-8986.2006.00392.x

Debaleena Chattopadhyay and Karl F. MacDorman. 2016. Familiar faces rendered strange: Why inconsistent realism drives characters into the uncanny valley (Sep. 2016). *Journal of Vision*, 16, 11:7, 1–25. https://doi.org/10.1167/16.11.7

Marcus Cheetham and Lutz Jäncke. 2013. Perceptual and category processing of the uncanny valley hypothesis' dimension of human likeness (Jun. 2013): Some methodological issues. *Journal of Visualized Experiments*, 76, 4375. https://doi.org/10.3791/4375

Marcus Cheetham, Ivana Pavlović, Nicola J. Jordan, Pascal Suter, and Lutz Jäncke. 2013. Category processing and the human likeness dimension of the uncanny valley hypothesis: Eye-tracking data. *Frontiers in Psychology*, 4, 108. https://doi.org/10.3389/fpsyg.2013.00108

Marcus Cheetham, Pascal Suter, and Lutz Jäncke. 2011. The human likeness dimension of the "uncanny valley hypothesis": Behavioral and functional MRI findings (Nov. 2011). *Frontiers in Human Neuroscience*, 5, 125, 126. https://doi.org/10.3389/fnhum.2011.00126

Marcus Cheetham, Pascal Suter, and Lutz Jäncke. 2014. Perceptual discrimination difficulty and familiarity in the uncanny valley: More like a "happy valley". *Frontiers in Psychology*, 5 (Nov. 2014), 1219. https://doi.org/10.3389/fpsyg.2014.01219

Marcus Cheetham, Lingdan D. Wu, Paul Pauli, and Lutz Jäncke. 2015. Arousal, valence, and the uncanny valley: Psychophysiological and self-report findings. *Frontiers in Psychology*, 6 (Jul. 2015), 981. https://doi.org/10.3389/fpsyg.2015.00981

Haiwen Chen, Richard Russell, Ken Nakayama, and Margaret Livingstone. 2010. Crossing the 'uncanny valley': Adaptation to cartoon faces can influence perception of human faces. *Perception*, 39, 3 (Aug. 2010), 378–386. https://doi.org/10.1068/p6492

Mike W.-L. Cheung. 2019. A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology Review, 29* (Aug. 2019), 387–396. https://doi.org/10.1007/s11065-019-09415-6

Leon Ciechanowski, Aleksandra Przegalińska, Mikolaj Magnuski, and Peter Gloor. 2019. In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, 92 (Mar. 2019), 539–548. https://doi.org/10.1016/j.future.2018.01.055

Jacob Cohen. 1988. *Statistical power analysis for the behavioral sciences* (2nd ed).. New Jersey: Lawrence Erlbaum Associates, Inc.

Valerie Curtis, Mícheál de Barra, and Robert Aunger. 2011. Disgust as an adaptive system for disease avoidance behavior. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366 (Feb. 2011), 389–401. https://doi.org/10.1098/rstb.2010.0117

Zhengyan Dai and Karl F. MacDorman. 2018. The doctor's digital double: How warmth, competence, and animation promote adherence intention. *PeerJ Computer Science,* 4 (2018), e168, 1–29. https://doi.org/10.7717/peerj-cs.168

Jason C. Deska, Steven M. Almaraz, and Kurt Hugenberg. 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 (Sep. 2016), 183–190. https://doi.org/10.1177/1948550616671404

Matthieu Destephe, Massimiliano Zecca, Kenji Hashimoto, and Atsuo Takanishi. 2014. Uncanny valley, robot and autism: Perception of the uncanniness in an emotional gait. *Proceedings of the IEEE International Conference on Robotics and Biomimetics* (pp. 1152–1157), Bali, Indonesia, 2014. https://doi.org/10.1109/ROBIO.2014.7090488.

Matthieu Destephe, Martim Brandao, Tatsuhiro Kishi, Massimiliano Zecca, Kenji Hashimoto, and Atsuo Takanishi. 2015. Walking in the uncanny valley: Importance of the attractiveness on the acceptance of a robot as a working partner. *Frontiers in Psychology*, 6 (Feb. 2015), 204. https://doi.org/10.3389/fpsyg.2015.00204

Alexander Diel and Karl F. MacDorman. 2021. Creepy cats and strange high houses: Support for configural processing in testing predictions of nine uncanny valley theories. *Journal of Vision*.

Shuyuan Feng, Xueqin Wang, Qiandong Wang, Jing Fang, Yaxue Wu, Li Yi, and Kunlin Wei. 2018. The uncanny valley effect in typically developing children and its absence in children with autism spectrum disorders. *PLoS ONE,* 13 (Nov. 2018), e0206343. https://doi.org/10.1371/journal.pone.0206343

Francesco Ferrari, Maria Paola Paladino, and Jolanda Jetten. 2016. Blurring human–machine distinctions: Anthropomorphic appearance in social robots as a threat to human distinctiveness. *International Journal of Social Robotics*, 8, 2 (Jan. 2016), 287–302. https://10.1007/s12369-016-0338-y

Anne E. Ferrey, Tyler J. Burleigh, and Mark J. Fenske. 2015. Stimulus-category competition, inhibition, and affective devaluation: A novel account of the uncanny valley. *Frontiers in Psychology*, 6 (Mar. 2015), 249. https://doi.org/10.3389/fpsyg.2015.00249

Susan T. Fiske, Amy J. C. Cuddy, & Peter Glick. 2007. Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11* (Feb. 2007), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Susan T. Fiske, Amy J. C. Cuddy, Peter Glick, & Jun Xu. 2002. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82 (Jun. 2002), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Rasmus Gahrn-Andersen. 2020. Seeming autonomy, technology and the uncanny valley. *AI & Society* (Aug. 2020). https://doi.org/10.1007/s00146-020-01040-9

Kurt Gray and Daniel M. Wegner. 2012. Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 125 (Oct. 2012), 125–130. https://doi.org/10.1016/j.cognition.2012.06.007

Robert D. Green, Karl F. MacDorman, Chin-Chang Ho, and Sandosh K. Vasudevan. 2008. Sensitivity to the proportions of faces that vary in human likeness. *Computers in Human Behavior*, 24, 5 (Sep. 2008), 2456–2474. https://doi.org/10.1016/j.chb.2008.02.019

Sigmund Freud. 1919/2003. The uncanny [das Unheimliche] (D. McClintock, Trans.). Penguin, New York.

Ismet Handzic and Kyle B. Reed. 2015. Perception of gait patterns that deviate from normal and symmetric biped locomotion. *Frontiers in Psychology*, 6 (Feb. 2015). https://doi.org/10.3389/fpsyg.2015.00199

David Hanson, Andrew Olney, Steve Prilliman, Eric Mathews, Marge Zielke, Derek Hammons, Raul Fernandez, and Harry E. Stephanou. 2005. Upending the uncanny valley. *Proceedings of the Twentieth National Conference on Artificial Intelligence* (Jan. 2005), 1728–1729. AAAI Press, Menlo Park, CA.

Russell Hardin. 2002. *Trust and trustworthiness*. New York: Russell Sage Foundation.

Chin-Chang Ho, and Karl F. MacDorman. 2010. Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, 26 (Nov. 2010), 1508–1518. https://doi.org/10.1016/j.chb.2010.05.015

Chin-Chang Ho and Karl F. MacDorman. 2017. Measuring the uncanny valley effect: Refinements to indices for perceived humanness, attractiveness, and eeriness. *International Journal of Social Robotics*, 9 (Jan. 2017), 129–139. https://doi.org/10.1007/s12369-016-0380-9

Chin-Chang Ho, Karl F. MacDorman, and Zacharias A. D. Pramono. 2008. Human emotion and the uncanny valley: A GLM, MDS, and ISOMAP analysis of robot video ratings. *Proceedings of the Third ACM/IEEE International Conference on Human-Robot Interaction* (Jan. 2008), pp. 169–176, March 11–14, 2008. Amsterdam, Netherlands. https://doi.org/10.1145/1349822.1349845

Kurt Hugenberg. 2005. Social categorization and the perception of facial affect: Target race moderates the response latency advantage for happy faces. *Emotion,* 5, 3, 267–276. https://doi.org/10.1037/1528-3542.5.3.267

Yoonhyuk Jung and Eunae Cho. 2018. Context-specific affective and cognitive responses to humanoid robots. *Proceedings of the 22nd ITS Biennial Conference, Beyond the Boundaries: Challenges for Business, Policy and Society* (Jun. 2018). International Telecommunications Society (ITS). Seoul, Korea.

Hiroko Kamide, Koji Kawabe, Satoshi Shigemi, and Tatsuo Arai. 2013. Development of a psychological scale for general impressions of humanoid. *Advanced Robotics,* 27, 1, 3–17, https://doi.org/10.1080/01691864.2013.751159

Jari Kätsyri. 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, 1362. https://doi.org/10.3389/fpsyg.2018.01362

Jari Kätsyri, Beatrice de Gelder, and Apio Takala. 2019. Virtual faces evoke only a weak uncanny valley effect: An empirical investigation with controlled virtual face images. *Perception*, 48, 10 (Aug. 2019), 968–991. https://doi.org/10.1177/0301006619869134

Jari Kätsyri, Klaus Förger, Meeri Mäkäräinen, and Tapio Takala. 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 (Apr. 2015), 390. https://doi.org/10.3389/fpsyg.2015.00390

Jari Kätsyri, Meeri Mäkäräinen, and Tapio Takala. 2017. Testing the 'uncanny valley' hypothesis in semirealistic computer-animated film characters: An empirical evaluation of natural film stimuli, *International Journal of Human-Computer Studies,* 97 (Jan. 2017), 149–161. https://doi.org/10.1016/j.ijhcs.2016.09.010.

Andrew Kennedy. 2014. *The effect of color on emotions in animated films*. Open Access Theses, 201 (Spring 2014). https://docs.lib.purdue.edu/open_access_theses/201

Marino Kimura and Yuko Yotsumoto. 2018. Auditory traits of "own voice." *PLOS One*, 13, 6 (Jun. 2016), Article e0199443. https://doi.org/10.1371/journal.pone.0199443

Kami Koldewyn, Patricia Hanus, and Benjamin Balas. 2014. Visual adaptation of the perception of "life": Animacy is a basic perceptual dimension of faces. *Psychonomic Bulletin and Review*, 21, 4 (2014), 969–975. https://doi.org/10.3758/s13423-013-0562-5

Genevieve M. Kozak, Megan L. Head, Alycia C. R. Lackey, and Janette W. Boughman. 2013. Sequential mate choice and sexual isolation in threespine stickleback species. *Journal of Evolutionary Biology*, 26 1 (Jan. 2013), 130–140. https://doi.org/10.1111/jeb.12034

Katharina Kühne, Martin H. Fischer, and Yuefang Zhou. 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

Oliver Langner, Ron Dotsch, Gijsbert Bijlstra, Daniel H. J. Wigboldus, Skyler T. Hawk, and Ad van Knippenberg. 2010. Presentation and validation of the Radboud Faces Database. *Cognition & Emotion,* 24, 8 (Nov. 2010), 1377—1388. https://doi.org/10.1080/02699930903485076

Markus Langer and Cornelius J. König. 2018. Introducing and testing the creepiness of situation scale (CRoSS). *Frontiers in Psychology*, 9 (Nov. 2018), 2220. https://doi.org/10.3389/fpsyg.2018.02220

Stephanie Lay, Nicola Brace, Graham Pike, and Frank Pollick. 2016. Circling around the uncanny valley: Design principles for research into the relation between human likeness and eeriness. *i-Perception,* 7(6), 1–11. https://doi.org/10.1177/2041669516681309

David J. Lewkowicz and Asif A. Ghazanfar. 2012. The development of the uncanny valley in infants. *Developmental Psychobiology,* 54, 2, 124–132. https://doi.org/10.1002/dev.20583

Chaolan Lin, Selma Šabanović, Lynn Dombrowski, Andrew D. Miller, Erin Brady and Karl F. MacDorman. 2021. Parental acceptance of children's storytelling robots: A projection of the uncanny valley of AI. *Frontiers in Robotics and AI*, 8 (May 2021), 579993, 1–15. https://doi.org/10.3389/frobt.2021.579993

Tanja Lischetzke, David Izydorczyk, Christina Hüller, and Markus Appel. 2017. The topography of the uncanny valley and individuals' need for structure: A nonlinear mixed effects analysis. *Journal of Research in Personality*, 68 (Jul. 2011), 96–113. https://doi.org/10.1016/j.jrp.2017.02.001

Lukasz Piwek, Lawrie S. McKay, and Frank E. Pollick. 2014. Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition,* 130, 3 (2014), 271–277. https://doi.org/10.1016/j.cognition.2013.11.001.

Diana Löffler, Judith Dörrenbächer, and Marc Hassenzahl. 2020. The uncanny valley effect in zoomorphic robots: The *U*-shaped relation between animal likeness and likeability. In *Proceedings of the 2020 ACM/IEEE International Conference on Human–Robot Interaction* (pp. 261–270). New York, NY: ACM. https://doi.org/10.1145/3319502.3374788

Christine E. Looser and Thalia Wheatley. 2010. The tipping point of animacy: How, when, and where we perceive life in a face. *Psychological Science*, 21, 12 (Dec. 2010), 1854–1862. https://doi.org/10.1177/0956797610388044

Paweł Łupkowski, Marek Rybka, Dagmara Dziedzic, and Wojciech Włodarczyk. 2019. The background context condition for the uncanny valley hypothesis. *International Journal of Social Robotics*, 11 (Sep. 2018), 25–33. https://doi.org/10.1007/s12369-018-0490-7

Goh Matsuda, Hiroshi Ishiguro, and Kazuo Hiraki. 2015. Infant discrimination of humanoid robots. *Frontiers in Psychology*, 6 (Sep. 2015), 1397. https://doi.org/10.3389/fpsyg.2015.01397

Yoshi-Taka Matsuda, Yoko Okamoto, Misako Ida, Kazuo Okanoya, and Masako Myowa-Yamakoshi. 2012. Infants prefer the faces of strangers or mothers to morphed faces: An uncanny valley between social novelty and familiarity. *Biology Letters*, 8 (Oct. 2012), 725–728. https://10.1098/rsbl.2012.0346

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

Karl F. MacDorman and Debaleena Chattopadhyay. 2017. Categorization-based stranger avoidance does not explain the uncanny valley. *Cognition*, 161 (Jan. 2017), 129–135. https://doi.org/10.1016/j.cognition.2017.01.009

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

Karl F. MacDorman, Robert D. Green, Chin-Chang Ho, and Clinton T. Koch. 2009. Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior*, 25, 3 (Dec. 2014), 695–710. https://doi.org/10.1016/j.chb.2008.12.026

Karl F. MacDorman and Hiroshi Ishiguro. 2006. The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, 7, 3 (Jan. 2006), 297–337. https://doi.org/10.1075/is.7.3.03mac

Karl F. MacDorman, Takashi Minato, Michihiro Shimada, Shoji Itakura, Stephen Cowley, and Hiroshi Ishiguro. 2005. Assessing human likeness by eye contact in an android testbed. *Proceedings of the XXVII Annual Meeting of the Cognitive Science Society* (Jul. 2005), pp. 1373–1378.

Karl F. MacDorman, Preethi Srinivas, and Himalaya Patel. 2013. The uncanny valley does not interfere with level 1 visual perspective taking. *Computers in Human Behavior*, 29, 4 (Jul. 2013), 1671–1685. https://doi.org/10.1016/j.chb.2013.01.051

Meeri Mäkäräinen, Jari Kätsyri, and Tapio Takala. 2014. Exaggerating facial expressions: A way to intensify emotion or a way to the uncanny valley? *Cognitive Computation*, 6, 4 (May 2014), 708–721. https://doi.org/10.1007/s12559-014-9273-0

Bruce Mangan. 2015. The uncanny valley as fringe experience. *Interaction Studies*, 16, 2 (Sep. 2015), 193–199. https://doi.org/10.1075/is.16.2.05man

Maya B. Mathur and David B. Reichling. 2016. Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. *Cognition*, 146 (Jan. 2016), 22–32. https://doi.org/10.1016/j.cognition.2015.09.008

Maya B. Mathur, David B. Reichling, Francesca Lunardini, Alice Geminiani, Alberto Antonietti, Peter A. M. Ruijten, Carmel A. Levitan, Gideon Nave, Dylan Mafredi, Brandy Bessette-Symons, Attila Szuts, and Balazs Aczel. 2020. Uncanny but not confusing: Multisite study of perceptual category confusion in the uncanny valley. *Computers in Human Behavior*, 103 (Feb. 2020), 21–30. https://doi.org/10.1016/j.chb.2019.08.029

Koh Matsuda, Hiroshi Ishiguro, and Kazuo Hiraki. 2015. Infant discrimination of humanoid robots. *Frontiers in Psychology, 6,* 1397. https://doi.org/10.3389/fpsyg.2015.01397

Yoshi-Taka Matsuda, Yoko Okamoto, Misako Ida, Kazuo Okanoya, and Masako Myowa-Yamakoshi. 2012. Infants prefer the faces of strangers or mothers to morphed faces: An uncanny valley between social novelty and familiarity. *Biology Letters*, 8, 5 (Oct. 2012), 725–728. https://doi.org/10.1098/rsbl.2012.0346

Francis T. McAndrew and Sara S. Koehnke. 2016. On the nature of creepiness. *New Ideas in Psychology*, 43 (Dec. 2016), 10–15. https://doi.org/10.1016/j.newideapsych.2016.03.003

Rachel McDonnell and Martin Breidt. 2010. Face reality: Investigating the uncanny valley for virtual faces. In Marie-Paule Cani and Alla Sheffer (Eds.), ACM SIGGRAPH Asia Sketches (Jan. 2010), pp. 1–2. ACM Press, New York, NY, USA.

Rachel McDonnell, Martin Breidt, M., and Heinrich H. Bülthoff. 2012. Render me real? Investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics*, 31 (Jul. 2012), 1–11. https://doi.org/10.1145/2185520.2185587

Lianne F. S. Meah and Roger K. Moore. 2014. The uncanny valley: A focus on misaligned cues. In Michael Beetz, Benjamin Johnston, & Mary-Anne Williams (Eds.), *Social Robotics: 6th*

International Conference (Oct. 2014), pp. 256–265. ICSR Proceedings. Sydney, NSW, Australia. October 27–29.

Wade J. Mitchell, Chin-Chang Ho, Himalaya Patel, and Karl F. MacDorman. 2011. Does social desirability bias favor humans? Explicit–implicit evaluations of synthesized speech support a new HCI model of impression management. *Computers in Human Behavior, 27*(1), 402–412. https://doi.org/10.1016/j.chb.2010.09.002

Wade J. Mitchell, Kevin A. Szerszen, Amy Shirong Lu, Paul W. Schermerhorn, Matthias Scheutz, and Karl F. MacDorman. 2011b. A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception*, 2, 1 (Mar. 2011), 10–12. https://doi.org/10.1068/i0415

Roger K. Moore. 2012. A Bayesian explanation of the 'uncanny valley' effect and related psychological phenomena. *Scientific Reports,* 2 (Nov. 2012), 864. https://doi.org/10.1038/srep00864

Mahdi Muhammad Moosa and S. M. Minhaz Ud-Dean. 2010. Danger avoidance: An evolutionary explanation of uncanny valley. Biology Theory, 5 (Apr. 2010), 12–14. https://doi.org/10.1162/BIOT_a_00016

Richard L. Moreland and Robert B. Zajonc. 1982. Exposure effects in person perception: Familiarity, similarity, and attraction. *Journal of Experimental Social Psychology*, 18, 5 (Dec. 1980), 395–415. https://doi.org/10.1016/0022-1031(82)90062-2

Masahiro Mori. 2012. The uncanny valley (Karl F. MacDorman & Norri Kageki, Trans.). *IEEE Robotics and Automation,* 19, 2 (Jun. 2012), 98–100. (Original work published in 1970). https://doi.org/10.1109/MRA.2012.2192811

Vicneas Muniady and Ahmad Zamzuri Mohamad Ali. 2020. The effect of valence and arousal on virtual agent's designs in quiz based multimedia learning environment. *International Journal of Instruction, 13*(4), 903-920. https://doi.org/10.29333/iji.2020.13455a

Hiroshi Nitta and Kazuhide Hashiya. 2021. Self-face perception in 12-month-old infants: A study using the morphing technique. *Infant Behavior and Development, 62* (Feb. 2021), 101479. https://doi.org/10.1016/j.infbeh.2020.101479

Iroju Olaronke, Oluwaseun A. Ojerinde, and Rhoda Ikono. 2017. State of the art: A study of human–robot interaction in healthcare. *International Journal of Information Engineering & Electronic Business*, 9, 3 (May 2017), 43–55. https://doi.org/10.5815/ijieeb.2017.03.06

Maike Paetzel, Christopher E. Peters, Ingela Nyström, and Ginevra Castellano. 2016. Effects of multimodal cues on children's perception of uncanniness in a social robot. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction* (Oct. 2016), pp. 297–301. Association for Computing Machinery. https://doi.org/10.1145/2993148.2993157

Jussi P. Palomäki, Anton Kunnari, Marianna Drosinou, Mika Koverola, Noora Lehtonen, Juho Halonen, Marko Repi, and Michael Laakasuo. 2018. Evaluating the replicability of the uncanny valley effect. *Heliyon,* 4, 11 (Nov. 2018). https://doi.org/10.1016/j.heliyon.2018.e00939

Christopher J. Patrick and Stacey A. Lavoro. 1997. Ratings of emotional response to pictorial stimuli: Positive and negative affect dimensions. *Motivation and Emotion*, 21 (Dec. 1997), 297–321. https://doi.org/10.1023/A:1024432322584

Andrea Paulus and Dirk Wentura. 2015. It depends: Approach and avoidance reactions to emotional expressions are influenced by the contrast emotions presented in the task. *Journal of Experimental Psychology: Human Perception and Performance*, 42, 2, 197–212. https://doi.org/10.1037/xhp0000130

Jaime Alvarez Perez, Hideki Garcia Goo, Ana Sánchez Ramos, Virginia Contreras, and Megan Strait. *Companion of the 2020 ACM/IEEE International Conference on Human–Robot Interaction* (pp. 101–103), March 2020. https://doi.org/10.1145/3371382.3378312

Lukasz Piwek, Lawrie S. McKay, and Frank E. Pollick. 2014. Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition, 130*(Mar. 2014), 271–277. https://doi.org/10.1016/j.cognition.2013.11.001

Ellen Poliakoff, Natalie Beach, Rebecca Best, Toby Howard, and Emma Gowen. 2013. Can looking at a hand make your skin crawl? Peering into the uncanny valley for hands. *Perception*, 42, 9 (Aug. 2015), 998–1000. https://doi.org/10.1068/p7569

Akanaksha Prakash and Wendy A. Rogers. 2015. Why some humanoid faces are perceived more positively than others: Effects of human-likeness and task. *International Journal of Social Robotics*, 7, 2, 309–331. https://doi.org/10.1007/s12369-014-0269-4

Connor P. Principe and Judith H. Langlois. 2011. Faces differing in attractiveness elicit corresponding affective responses. *Cognition & Emotion*, 25, 1 (2011), 140–148. https://doi.org/10.1080/02699931003612098

Si Qiao and Roger Eglin. 2011. Accurate behaviour and believability of computer generated images of human head. *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry* (pp. 545–548), December 2011. https://doi.org/10.1145/2087756.2087860

Si Qiao, Roger Eglin, and Ariel Beck. 2011. Audience perception of computer generated human facial behaviour. *GSTF International Journal on Computing*, 1, 3 (April 2011), 61–65.

Christopher H. Ramey. 2005. The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. In *Proceedings of Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots* (Dec. 2005), pp. 8–13. Tsukuba, Japan.

Alexandra S. Rativa, Marie Postma, and Menno van Zaanen. 2019. The uncanny valley of the virtual (animal) robot. In Munir Merdan, Wilfried Lepuschitz, Gottfried Koppensteiner, Richard Balogh, and David Obdržálek (Eds.), *Robotics in Education. RiE 2019*. Advances in Intelligent Systems and Computing, vol. 1023. Springer, Cham. https://doi.org/10.1007/978-3-030-26945-6_38

Josh D. Redstone. 2013. Beyond the uncanny valley: A theory of eeriness for android science research. Master's thesis. https://doi.org/10.22215/etd/2013-09987

Jasia Reichardt. 1978. Human reactions to imitation humans, or Masahiro Mori‘s uncanny valley. In Jasia Reichardt, *Robots: Fact, fiction, and prediction* (1st ed., pp. 26–27). Viking, New York.

Anne Reuten, Maureen van Dam, and Marnix Naber. 2018. Pupillary responses to robotic and human emotions: The uncanny valley and media equation confirmed. *Frontiers in Psychology*, 23, 9 (Mar. 2018), 774. https://doi.org/10.3389/fpsyg.2018.00774

Astrid M. Rosenthal-von der Pütten and Nicole C. Krämer. 2014. How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior*, 36 (Jul. 2014), 422–439. https://doi.org/10.1016/j.chb.2014.03.066

Astrid M. Rosenthal-von der Pütten, Nicole Krämer, Stefan Maderwald, Matthias Brand, and Fabian Grabenhorst. 2019. Neural mechanisms for accepting and rejecting artificial social partners in the uncanny valley. *The Journal of Neuroscience*, 39, 33 (Aug. 2019), 6555–6570. https://doi.org/10.1523/JNEUROSCI.2956-18.2019

Nicholas Royle. 2003. *The Uncanny: An Introduction*. Manchester University Press, New York.

Stefania Sansoni, Andrew Wodehouse, Angus K. McFadyen, and Arjan Buis. 2015. The aesthetic appeal of prosthetic limbs and the uncanny valley: The role of personal characteristics in attraction. *International Journal of Design,* 9, 67–81.

Kyoshiro Sasaki, Keiko Ihaya, and Yuki Yamada. 2017. Avoidance of novelty contributes to the uncanny valley. *Frontiers in Psychology*, 8 (Mar. 2018), 1792. https://doi.org/10.3389/fpsyg.2017.01792

Ayse Pinar Saygin, Thierry Chaminade, Hiroshi Ishiguro, Jon Driver, and Chris Frith. 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 (Apr. 2011), 413–422. https://doi.org/10.1093/scan/nsr025

Sebastian Schindler, Eduard Zell, Mario Botsch, and Johanna Kissler. 2017. Differential effects of face-realism and emotion on event-related brain potentials and their implications for the uncanny valley theory. *Scientific Reports*, 7 (Mar. 2017), 45003. https://doi.org/10.1038/srep45003

Edward Schneider, Yifan Wang, and Shanshan Yang. 2009. Exploring the uncanny valley with Japanese video game characters. In B. Akira (Ed.), *Proceedings of the Digital Games Research Association (DiGRA): Situated Play* (Oct. 2017), pp. 546–549.

Jordan Schoenherr and Tyler J. Burleigh. 2015. Uncanny sociocultural categories. *Frontiers in Psychology,* 5 (Jan. 2015), 1456. https://doi.org/10.3389/fpsyg.2014.01456

Valentin Schwind, Pascal Knierim, Cagri Tasci, Patrick Franczak, Nico Haas, and Niels Henze. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, May 2017, Pages 1577–1582. https://doi.org/10.1145/3025453.3025602

Valentin Schwind, Katharina Leicht, Solveigh Jäger, Katrin Wolf, and Niels Henze. 2018. Is there an uncanny valley of virtual animals? A quantitative and qualitative investigation. *International Journal of Human-Computer Studies*, 111 (Mar. 2018), 49–61. https://doi.org/10.1016/j.ijhcs.2017.11.003

Jun'ichiro Seyama and Ruth S. Nagayama. 2007. The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments*, 16 (Aug. 2007), 337–351. https://doi.org/10.1162/pres.16.4.337

Mincheol Shin, Se Jung Kim, and Frank Biocca. 2019. The uncanny valley: No need for any further judgments when an avatar looks eerie. *Computers in Human Behavior,* 94 (May 2019), 100–109. https://doi.org/10.1016/j.chb.2019.01.016

Mincheol Shin, Stephen W. Song, and Tamara M. Chock. 2019. Uncanny valley effects on friendship decisions in virtual social networking service. *Cyberpsychology, Behavior, and Social Networking*. Advance online publication (Nov. 2019). https://doi.org/10.1089/cyber.2019.0122

Jacqueline C. Snow, Rafal M. Skiba, Taylor L. Coleman, and Marian E. Berryhill. 2014. Real-world objects are more memorable than photographs of objects. *Frontiers in Human Neuroscience,* 8, 837. https://doi.org/10.3389/fnhum.2014.00837

Shawn A. Steckenfinger and Asif A. Ghazanfar. 2009. Monkey visual behavior falls into the uncanny valley. *Proceedings of the National Academy of Sciences of the United States of America* (PNAS), 106, 43 (Oct. 2009), 18362–18366. https://doi.org/10.1073/pnas.0910063106

Jan-Philipp Stein and Peter Ohler. 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 (Mar. 2017), 43–50. https://doi.org/10.1016/j.cognition.2016.12.010

Jan-Philipp Stein and Peter Ohler. 2018. Uncanny...but convincing? Inconsistency between a virtual agent's facial proportions and vocal realism reduces its credibility and attractiveness, but not its persuasive success. *Interacting With Computers*, *30* (Nov. 2018), 480–491. https://doi.org/10.1093/iwc/iwy023

Megan K. Strait, Victoria A. Floerke, Wendy Ju, Keith Maddox, Jessica D. Remédios, Malte F. Jung, and Heather L. Urry. 2017. Understanding the uncanny: Both atypical features and category ambiguity provoke aversion toward humanlike robots. *Frontiers in Psychology*, 8 (Aug. 2017), 1366. https://doi.org/10.3389/fpsyg.2017.01366

Megan Strait and Matthias Scheutz. 2014. Measuring users' responses to humans, robots, and human-like robots with functional near infrared spectroscopy. The *23rd IEEE International Symposium on Robot and Human Interactive Communication* (Aug. 2014), 1128–1133. https://doi.org/10.1145/2702123.2702415

Megan Strait, M., Lara Vujovic, Victoria Floerke, Matthias Scheutz, and Heather L. Urry. 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* (Apr. 2015), 3593–3602. Seoul, Republic of Korea. https://doi.org/10.1145/2702123.2702415

Megan Strait, Heather L. Urry, and Paul Muentener. 2019. Children's responding to humanlike agents reflects an uncanny valley. In *Proceedings of the 14th ACM/IEEE International Conference on Human–Robot Interaction* (Mar. 2019), pp. 506–515. https://doi.org/10.1109/HRI.2019.8673088

Kohske Takahashi, Haruaki Fukuda, Kazuyuki Samejima, Katsumi Watanabe, and Kazuhiro Ueda. 2015. Impact of stimulus uncanniness on speeded response. *Frontiers in Psychology*, 6 (May 2015), 662. https://doi.org/10.3389/fpsyg.2015.00662

James C. Thompson, J. Gregory Trafton, and Patrick McKnight. 2011. The perception of humanness from the movements of synthetic agents. *Perception*, 40, 6 (Jan. 2011), 695–704. https://doi.org/10.1068/p6900

Angela Tinwell. 2009. Uncanny as usability obstacle. In A. Ant Ozok and Panayiotis Zaphiris (Eds.), *Online Communities and Social Computing*. Lecture Notes in Computer Science, vol. 5621 (Jul. 2009). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-02774-1_67

Angela Tinwell and Mark N. Grimshaw. 2009. Survival horror games—An uncanny modality. *Thinking After Dark,* 23 (Apr. 2009). Retrieved from: http://ubir.bolton.ac.uk/id/eprint/235

Angela Tinwell, Mark N. Grimshaw, and Deborah A. Nabi. 2015. The effect of onset asynchrony in audio-visual speech and the uncanny valley in virtual characters. *International Journal of Mechanisms and Robotic Systems*, 2, 2 (Apr. 2015), 97–110. https://doi.org/10.1504/IJMRS.2015.068991

Angela Tinwell, Mark N. Grimshaw, and Deborah A. Nabi. 2014. The uncanny valley and nonverbal communication in virtual characters. In Theresa Jean Tanenbaum, Magy Seif el-Nasr, & Michael Nixon (Eds.), *Nonverbal Communication in Virtual Worlds: Understanding and Designing Expressive Characters* (Jan. 2014), pp. 325–341. Carnegie Mellon University Press, Pittsburgh, PA.

Angela Tinwell, Mark N. Grimshaw, Deborah A. Nabi, and Andrew Williams. 2011. Facial expression of emotion and perception of the uncanny valley in virtual characters. *Computers in Human Behavior*, 2 (Nov. 2010), 741–749. https://doi.org/10.1016/j.chb.2010.10.018

Angela Tinwell, Deborah A. Nabi, and John P. Charlton. 2013. Perception of psychopathy and the uncanny valley in virtual characters. *Computers in Human Behavior*, 29, 4 (Mar. 2013), 1617–1625. https://doi.org/10.1016/j.chb.2013.01.008

Angela Tinwell and Robin J. S. Sloan. 2014. Children's perception of uncanny human-like virtual characters. *Computers in Human Behavior*, 36 (May 2014), 286–296. https://doi.org/10.1016/j.chb.2014.03.073

Fangwu Tung. 2016. Child perception of humanoid robot appearance and behavior. *International Journal of Human–Computer Interaction*, 32 (Apr. 2016), 493–502. https://doi.org/10.1080/10447318.2016.1172808

Burcu A. Urgen, Marta Kutas, & Ayse P. Saygin. 2018. Uncanny valley as a window into predictive processing in the social brain. *Neuropsychologia*, 114, 181–185. https://doi.org/10.1016/j.neuropsychologia.2018.04.027

Patricia Valdez and Albert Mehrabian. 1994. Effects of color on emotions. *Journal of Experimental Psychology: General*, 123, 4 (Jul. 2015), 394–409. https://doi.org/10.1037/0096-3445.123.4.394

Wolfgang Viechtbauer and Mike W.-L. Cheung. 2010. Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods,* 1, 2 (April/June 2010), 112–25.

Shensheng Wang, Scott O. Lilienfeld, and Philippe Rochat. 2015. The uncanny valley: Existence and explanations. *Review of General Psychology*, 19 (Dec. 2015), 393–407. https://doi.org/10.1037/gpr0000056

Shensheng Wang and Philippe Rochat. 2017. Human perception of animacy in light of the uncanny valley phenomenon. *Perception*, 46, 12 (Dec. 2017), 1386–1411. https://doi.org/10.1177/0301006617722742

Shensheng Wang, Yuk F. Cheong, Daniel D. Dilks. and Philippe Rochat. 2020. The uncanny valley phenomenon and the temporal dynamics of face animacy perception. *Perception*, 49(2020), 1069–1089. https://doi.org/10.1177/0301006620952611

Patrick P. Weis and Eva Wiese. 2017. Cognitive conflict as possible origin of the uncanny valley. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61 (Sep. 2017), 1599–1603. https://doi.org/10.1177/1541931213601763

Megan T. Wyman, Benjamin D. Charlton, Yann Locatelli, and David Reby. 2011. Variability of female responses to conspecific vs. heterospecific male mating calls in polygynous deer: An open door to hybridization? *PLOS One*, 6, 8 (Aug. 2011). https://doi.org/10.1371/journal.pone.0023296

Yuki Yamada, Takahiro Kawabe, and Keiko Ihaya. 2013. Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *Japanese Psychological Research*, 55, 1 (Aug. 2011), 20–32. https://doi.org/10.1111/j.1468-5884.2012.00538.x

Joachim von Zitzewitz, Patrick M. Boesch, Peter Wolf, and Robert Riener. 2013. Quantifying the human likeness of a humanoid robot. *International Journal of Social Robotics*, 5 (Jan. 2013), 263–276. https://doi.org/10.1007/s12369-012-0177-4

Angela Tinwell. 2009. Uncanny as usability obstacle. In A. Ant Ozok and Panayiotis Zaphiris (Eds.), *Online Communities and Social Computing*. Lecture Notes in Computer Science, vol. 5621 (Jul. 2009). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-02774-1_67

Eduard Zell, Carlos Aliaga, Adrian Jarabo, Katja Zibrek, Diego Gutierrez, Rachel McDonnell, and Mario Botsch. 2015. To stylize or not to stylize? The effect of shape and material stylization on the perception of computer-generated faces. *ACM Transactions on Graphics,* 34, 6 (Nov., 2015), 184, 1–12. https://doi.org/10.1145/2816795.2818126

Jie Zhang, Shuo Li, Jing-Yu Zhang, Feng Du, Yue Qi, and Xun Liu. 2020. A literature review of the research on the uncanny valley. In Rau, P. L. (Ed.), *Cross-Cultural Design: User Experience of Products, Services, and Intelligent Environments*. Lecture Notes in Computer Science, vol. 12192 (July 2020). Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-030-49788-0_19

Jakub A. Złotowski, Hidenobu Sumioka, Shuichi Nishio, Dylan F. Glas, Christoph Bartneck, and Hiroshi Ishiguro. 2015. Persistence of the uncanny valley: The influence of repeated interactions and a robot's attitude on its perception. *Frontiers in Psychology*, 6 (Jun. 2015), 883. https://doi.org/10.3389/fpsyg.2015.00883

A. APPENDIX

Table A1. Indices and Cronbach's α's of UV studies.

Authors (year) [study no.]	Indices: separate scales	UV effect significance?	Cronbach's α per condition	Stimulus creation technique
Bartneck et al. (2009a)	Likability: awful–nice, unfriendly–friendly, unkind–kind, and unpleasant–pleasant	No	.92, .88, .84	Real-life encounter
Destephe et al. (2015)	Eeriness: eerie–reassuring, freaky–numbing, supernatural–ordinary, spine-tingling–uninspiring, thrilling–boring, mortal–predictable, uncanny–bland, and hair-raising–unemotional	Yes	.85	Motion manipulation
Ho & MacDorman (2017)	Eeriness: dull–freaky, predictable–eerie, plain–weird, ordinary–supernatural, boring–shocking, uninspiring–spine-tingling, predictable–thrilling, bland–uncanny, and unemotional–hair-raising	Yes	.86	Distinct entities
Ho & MacDorman (2010)	Eeriness: reassuring–eerie, numbing–freaky, ordinary–supernatural, and uninspiring–spine-tingling	Yes	.74	Distinct entities
	Warmth: cold-hearted–warm-hearted, hostile–friendly, spiteful–well-intentioned, ill-tempered–good-natured, and grumpy–cheerful	Yes	.88	
Kätsyri, Mäkäräinen, & Takala (2017)	Likable: likable, aesthetic, and pleasant	No	.90	Distinct entities
	Eerie: eerie and unsettling	No	.70	

	Eerie: eerie, unsettling, and strange	No	.64	
Lischetzke et al. (2017)	Index: creepy, eerie, and uncanny	Yes	.92	Morphing
MacDorman & Chattopadhyay (2016)	Eeriness: ordinary–creepy, plain–weird, and predictable–eerie	No	N.A.	Realism render
	Warmth: cold-hearted–warm-hearted, hostile–friendly, and grumpy–cheerful	No	N.A.	
Mitchell et al. (2011b)	Eeriness (see Ho & MacDorman, 2010)	Yes	.70	Visuo-auditory mismatch
	Warmth (see Ho & MacDorman, 2010)	Yes	.88	
Rosenthal–von der Pütten & Krämer (2014)	Threatening: threatening, eerie, uncanny, dominant, and harmless	Maybe	.89	Distinct entities
	Likable: pleasant, likable, attractive, familiar, natural, and intelligent	Maybe	.83	
	Submissive: incompetent, weak, and submissive	No	.66	
	Unfamiliar: strange and unfamiliar	No	.67	
Schwind et al. (2018)	Familiarity: uncanny–familiar and freaky–numbing	Yes	N.A.	Distinct entities (cats)
	Aesthetics: ugly–beautiful and unaesthetic–aesthetic	Yes	N.A.	
Shin, Kim, & Biocca (2019)	Eeriness: reassuring–eerie, numbing–freaky, and ordinary–supernatural	Yes	.76	Realism render
Stein & Ohler (2018)	Eeriness (n.a.)	Yes	.74	Emotion manipulation, face distortion, realism render, visuo-auditory mismatch

Tinwell et al. (2013)	Uncanniness: eerie, nonhumanlike, repulsive, unattractive, unlikable, and unresponsive	Yes	.74, .80, .80	Emotion manipulation
Tung (2016) [1][2]	Social attraction: friendly, likable, and pleasant	Yes [1] No [2]	≥ .70	Distinct entities
Zlotowski et al. (2015)	Eeriness (n.a.)	Yes	.62 (lowest of three measurements)	Real-life encounter

Note. Eeriness and *Warmth* denote the indices developed by Ho and MacDorman (2010, 2017) and their derivations. We did not find studies with information on correlations between individual scale items.

Table A2. Summary and evaluation of stimulus creation techniques.

Stimulus creation technique	Exemplar studies	Advantages	Disadvantages	Further considerations
Distinct entities	Mathur et al., 2020 Rosenthal–von der Pütten & Krämer, 2014	Relatively high ecological validity, variable stimulus control, easy access	Confounding variables, no gradual range	Additional control when selecting stimuli can decrease confounding variables
Emotion manipulation	Tinwell et al., 2014	Specific, controllable stimulus manipulation	stimulus noise	
Face distortion	Mäkäräinen et al., 2014 MacDorman et al., (2009)	Controllable stimulus manipulation, gradual range	Stimulus noise	Strength of distortion should have a sufficient range

Morphing	Lischetzke et al., 2017 Sasaki, Ihaya, & Yamada	Controllable stimulus manipulation, gradual range	Results depend on endpoint stimuli choice, stimulus noise	Endpoint stimuli should be sufficiently distinct
Mismatch	Seyama & Nagayama, 2007	Controllable stimulus manipulation	Stimulus noise, no gradual range	Selection of mismatched features (e.g., eyes) Lack of research
Motion manipulation	Handzic & Reed, 2015			Lack of research
Realism render	McDonnell et al., 2012 MacDorman & Chattopadhyay, 2017	Controllable stimulus manipulation	Stimulus noise	
Real-life encounter	Zlotowski et al., 2015 Bartneck, Kanda, Ishiguro, & Hagita, 2009	High ecological validity for android science	Low internal validity, difficult setup and stimulus acquisition	Android/robotic and human counterpart stimuli should match Lack of research
Visuo-auditory mismatch	Mitchell et al., 2011b			Lack of research
Voice distortion	Baird et al.,2018			Lack of research