Lischetzke, T., Izydorczyk, D., Hüller, C., & Appel, M. (in press). The topography of the uncanny valley and individuals' need for structure: A nonlinear mixed effects analysis. *Journal of Research in Personality*. doi: 10.1016/j.jrp.2017.02.001

The Topography of the Uncanny Valley and Individuals' Need for Structure: A Nonlinear Mixed Effects Analysis

Tanja Lischetzke

David Izydorczyk

Christina Hüller

Markus Appel

University of Koblenz-Landau

Tanja Lischetzke, Department of Psychology, University of Koblenz-Landau, Germany; David Izydorczyk, Department of Psychology, University of Koblenz-Landau, Germany. Christina Hüller, Department of Psychology, University of Koblenz-Landau, Germany. Markus Appel, Department of Psychology, University of Koblenz-Landau, Germany.

Correspondence concerning the article should be addressed to Tanja Lischetzke, Department of Psychology, University of Koblenz-Landau, Fortstr. 7, 76829 Landau, Germany. E-mail: lischetzke@uni-landau.de

Abstract

The uncanny valley hypothesis suggests that robots that closely resemble humans elicit feelings of eeriness. We focused on individual differences in the uncanny valley experience, which have been largely neglected to date. Using a mixed effects modelling approach, we tested whether individual differences in the need for structure predict uncanny valley sensitivity. Two experiments (*Ns* = 226 and 336) with morphed stimuli confirmed the uncanny valley effect. A moderator effect of need for structure was found in Experiment 2, which used a fine-grained manipulation of human likeness, but not in Experiment 1, which used a 3-step manipulation. The results provide tentative evidence that individuals who respond negatively to a lack of structure show a more pronounced ("deeper") uncanny valley effect.

Keywords: uncanny valley; human-robot interaction; androids; eeriness; need for structure

The Topography of the Uncanny Valley and Individuals' Need for Structure: A Nonlinear Mixed Effects Analysis

Humanlike machines are a steady feature in works of fiction, ranging from the female automat Olympia in Hoffman's early 19th century story *The Sandman* to the replicants in the movie *Blade Runner* (Scott, 1982) and the synthetic men and women in the TV series *Humans* (Channel 4 / AMC) or *Westworld* (HBO). In recent years, humanoid robots have begun to leave the fictional world and enter real life. The production of robots is on the rise (International Federation of Robotics, 2015) with the introduction of robots that are meant to operate in people's everyday surroundings, do household chores, assist the elderly, or serve as a companion (e.g., Gates, 2007; Hurst, 2014; Wingfield, 2015). In recent years, roboticists have created robots with physical attributes that resemble humans. Such attributes are meant to facilitate the robots' performance in human environments (e.g., Hirai, Hirose, Haikawa, & Takenaka, 1998) and ease humans' interactions with the robots (e.g., MacDorman & Ishiguro, 2006). However, users' reactions to these robots have not always been positive. This new generation of robots has at times elicited strong reactions of discomfort and repulsion ("This thing is what nightmares get nightmares from," user comment in response to the video of a Japanese communications android uploaded to YouTube; cf. Mara, 2014).

Feelings of discomfort in response to humanlike robots have been examined within the framework of the uncanny valley hypothesis, which has attracted the widespread attention of scholars as well as journalists and the general public (e.g., Azarian, 2015; Eveleth, 2013; Jonette, 2016; Meth, 2014). The uncanny valley hypothesis (Mori, 1970 / Mori, MacDorman, & Kageki, 2012) suggests that the acceptance of a robot (or an artifact more generally) increases as the extent to which it resembles a human increases until a certain point of human likeness is reached. At this point, acceptance drops, and the robot is perceived to be maximally eerie. With even greater human likeness, the eeriness is reduced, and the robot is evaluated more positively. The segment of the theoretical human likeness continuum where eeriness dominates is called the uncanny valley. Despite the popularity of the concept in psychology and robotics, there is still a debate about whether this phenomenon exists at all (for an overview, see Wang, Lilienfeld, & Rochat, 2015). Our research extends the current literature in several ways: We examined users' eeriness responses not only as a function of robotic stimuli but as a function of robotic stimuli and users' individual differences. Theory suggests that individual differences in the preference for simple cognitive structures (need for structure; Neuberg & Newsom, 1993) predict the shape of the negative responses that represent the uncanny valley. Our analyses were based on a *mixed effects approach* (Judd, Westfall, & Kenny, 2012), which enabled us not only to add information about more or less susceptible individuals but also to increase the precision with which the effects of the experimental treatment could be estimated (Judd et al., 2012). As the uncanny valley hypothesis suggests a nonlinear relationship between human likeness and eeriness, linear as well as nonlinear relationships were examined.

1.1 Humanoid Robots and the Uncanny Valley

The uncanny valley hypothesis was first introduced in an exclusively theoretical essay by the Japanese robot scientist Masahiro Mori in 1970 (Mori, 1970 / Mori et al., 2012). Mori postulated a nonlinear relationship between the human likeness of an artificial entity (e.g., robots, prosthetic hands, puppets, zombies) and emotional reactions to this object. Mori suggested that as the human likeness of an object increases, users' perceptions of the robot become more positive. At a point of high—but not perfect—human likeness, people's affinity for the entity will drop and give way to an uncanny, eerie feeling. When the human likeness of this robot increases to a point where the robot is indistinguishable from a human being, the eeriness will disappear and positive emotions will dominate again (see Figure 1). The steep dip into eeriness is known as the "uncanny valley" (Mori also hypothesized that the shape of the graph changes, and the extrema are amplified when the object is moving, e.g., when a prosthetic hand starts moving, but see Piwek, McKay, & Pollick, 2014, who found no support for this prediction).

In the last decade in which the development of humanoid robots has gained momentum, researchers' interest in the uncanny valley hypothesis has increased remarkably (cf. Kätsyri, Förger, Mäkäräinen, & Takala, 2015; Wang et al., 2015). In one of the now classic early studies on the uncanny valley, MacDorman and Ishiguru (2006) presented images of a mechanical looking humanoid, a human, and a series of morphed images that blended the robot and the human to varying degrees. They observed an increase in eeriness for the morphed images, which appeared to be consistent with the uncanny valley hypothesis (no statistical test for linear or curvilinear relationships was conducted). However, in the years that followed, several studies failed to detect a relationship that resembled the uncanny valley (Bartneck, Kanda, Ishiguro, & Hagita, 2007; Hanson, 2005; MacDorman, 2006; Poliakoff, Beach, Best, Howard, & Gowen, 2013; Seyama & Nagayama, 2007, cf. Kätsyri et al., 2015; Wang et al., 2015).

Some of these rather inconclusive results may be due to disagreements about how to operationalize the uncanny valley hypothesis. Many studies followed a morphing paradigm, but the stimuli used in the morphing process varied, including robots, androids (i.e., robots that are intended to mimic human beings as realistically as possible), and real humans (MacDorman & Ishiguro, 2006), dolls and real humans (Seyama & Nagayama, 2007), computer-generated avatars and real humans (Cheetham, Suter, & Jäncke, 2011), or only computer-generated faces (Burleigh, Schoenherr, & Lacroix, 2013.). Often the main dependent variable directly addressed feelings of eeriness (e.g., Burleigh et al., 2013; MacDorman, Green, Ho, & Koch, 2009; Wang, 2014), but other research focused on familiarity (Hanson, 2005, 2006; Tinwell, Grimshaw, & Williams, 2010), likeability (Bartneck et al., 2007; Yamada, Kawabe, & Ihaya, 2013), or pleasantness

(Seyama & Nagayama, 2007). The analyses typically lacked inferential statistics, and more specifically, (multilevel) polynomial regression techniques were rarely applied to test the nonlinearity assumption. It is important to note that variations between participants had rarely been taken into account. Individual differences are not only a theoretically and practically important source of variance (e.g., the question of which users are particularly sensitive to the uncanny valley phenomenon), but ignoring variations in participants' responses for a given stimulus can lead to biased estimates of the stimulus effects (Judd et al., 2012, see more below).

In one of the few studies that discussed individual differences, Moore (2012) proposed a specific weighting factor that reflected the sensitivity of an observer to any perceived perceptual conflict in his Bayesian model. However, he did not specify what this weighting factor could be. MacDorman and Entezari (2015) connected several possible constructs (perfectionism, neuroticism, anxiety, personal distress, animal reminder sensitivity, human-robot uniqueness, android-robot uniqueness, negative attitudes towards robots, and religious fundamentalism) to eeriness and warmth ratings in response to humanlike robots (androids with nonhuman features). They found positive relationships between all of the individual difference variables and eeriness (except for android-robot uniqueness), but it remained unclear how specific these relationships were to the stimuli that were reported on—that is, whether a human, a less humanlike robot, or both would yield similar relationships.

1.2 Explaining the Uncanny Valley Experience

Theoretical accounts backing the uncanny valley hypothesis date back to German psychologist Ernst Jentsch (1906/1997) who discussed uncanny sensations in response to humanlike (pseudo-) automata (e.g., the original mechanical Turk and examples from fiction). Jentsch suggested that the uncanny (*das Unheimliche*) might be caused by uncertainty about which category the entity belongs to, and that such uncertainty in turn might elicit a lack of orientation and a feeling of estrangement (Jentsch, 1906/1997, p. 217). Some empirical evidence has indicated that category uncertainty might indeed be a plausible mechanism for the uncanny valley hypothesis. Researchers identified a negative correlation between reaction times in a categorization task and the likeability ratings of artificial agents (Yamada et al., 2013). This relationship was found for both the avatar-to-human continuum and the avatar-to-dog continuum, but no such effect was found for mergers within the "human" category (i.e., stepwise visual transitions from one person's photo portrayal into another person's portrayal). As part of a recent series of studies, individuals were asked to rate the mechanical or human resemblance of a face (Mathur & Reichling, 2016). The time taken to make this judgment was greatest for faces that exhibited low likeability (but rating time did not serve as a statistical mediator).

A second theoretical account focuses on *expectancy violation* and human *prediction errors*. Several scholars have suggested that when a robot has a very humanlike appearance, it will activate a human schema. Due to its imperfections, however, the robot will eventually fail to measure up to these expectations. When the predictions that come along with the activation of the human model are violated, the artificial agent is then regarded as uncanny (MacDorman, 2006; Matsui, Minato, MacDorman, & Ishiguro, 2005; Mitchell, Szerszen, Lu, Schermerhorn, Scheutz, & MacDorman, 2011; Saygin, Chaminade, Ishiguro, Driver, & Frith, 2012; Steckenfinger & Ghazanfar, 2009). MacDorman and Chattopadhyay (2016) showed morphs between photographed faces and computer-generated faces and found support for the role of realism inconsistency which is assumed to evoke prediction errors, which in turn lead to aversive responses (however, they found no evidence for category uncertainty).

Gray and Wegner (2012) specified the incompatibility that leads to eerie feelings. On the basis of a general theory of mind perception (Gray, Gray, & Wegner, 2007), they found that robots that are associated with experience (i.e., the capacity to feel and to sense) elicit uncanny feelings. Agency (i.e., the capacity to engage in planned action and to exert self-control) was unrelated to feelings of eeriness, because unlike experience, agency fits the established mental model of robots.

1.3 Individual Differences in the Need for Structure and the Uncanny Valley

The approaches outlined above rest on the assumption that individuals are inclined to use simplified generalizations of previous experiences (e.g., schemata, prototypes, scripts) to process incoming information. The creation and use of these structures enables humans to interpret the world on the basis of limited cognitive resources. Although all healthy humans rely on these structures, there are meaningful differences in the extent to which individuals “are dispositionally motivated to cognitively structure their worlds in simple, unambiguous ways” (Neuberg & Newsom, 1993, p.114). Neuberg and Newsom’s *personal need for structure* (PNS) is closely related to constructs such as *intolerance of ambiguity* (Frenkel-Brunswik, 1949) or the *need for closure* (Kruglanski, Webster, & Klem, 1993). People with a high need for structure are supposed to create relatively homogeneous, well-bounded, and distinct cognitive structures, they perceive ambiguity and grey areas to be problematic and annoying, and they are expected to experience discomfort when structure and clarity appear to be missing (Thompson, Naccarato, Parker, & Moskowitz, 2001). A self-report scale is available to assess individual differences in PNS (Newson & Neuberg, 1993), which consists of two subdimensions, representing the desire to live a well-structured life (desire for structure; DFS) and the negative response to unstructured, unpredictable stimuli and situations (response to lack of structure; RLS).

People with higher levels of PNS were found to be more likely to use less complex representations in their trait-based representations of the elderly, to apply previously acquired social categories to new situations, and to use simpler ways of organizing social and nonsocial information (Neuberg & Newsom, 1993). Moreover, people high on PNS were more likely to form spontaneous trait inferences—one way to impose structure in social actions (Moskowitz, 1993). Furthermore, studies have shown that high PNS individuals have a greater tendency to develop and rely on stereotypes (as another form of simple cognitive structure) in ambiguous situations (Bar-Tal & Guinote, 2002; Clow & Esses, 2005; Neuberg & Newsom, 1993). Individuals high on PNS were found to be more cognitively rigid (less creative) when confronted with schema-inconsistent stimuli (Gocłowska, Baas, Crisp, & De Dreu, 2014), and they showed intense negative responses to meaning threats (Landau et al., 2004; McGregor, Haji, & Kang, 2008; cf. Heine, Proulx, & Vohs, 2006).

The available research on the two subdimensions has suggested qualitative differences. The RLS factor was associated with both trait and social anxiety (whereas the DFS factor was not; Neuberg & Newsom, 1993) and DFS, but not RLS, was related to conscientiousness. More recent research has suggested that DFS is associated with sensitivity to the positive emotional qualities of stimuli, whereas RLS is associated with sensitivity to the negative emotional qualities of stimuli (Cavazos, Judice-Campbell, & Ditzfeld, 2012). In sum, individuals high on PNS tend to create and use simple, well-bounded, and distinct cognitive structures, they react particularly aversively to a lack of structure and clarity (Thompson et al., 2001), and their responses are more negative when they feel that structure is threatened (Cavazos et al., 2012), with the subdimension RLS showing particularly strong relationships with negative responses whenever structure is disrupted.

We assume that encountering a humanlike robot constitutes a situation in which expected generalizations of previous experiences (e.g., schemata) do not provide sufficient orientation. This should be particularly troubling for individuals who are dispositionally inclined to rely on these generalizations, that is, individuals high on PNS and particularly those with strong responses to a lack of structure (RLS). Thus, we expected that the shape of the uncanny valley would systematically vary with individuals' RLS, with a deeper valley for individuals high on RLS and a less deep valley for individuals low on RLS.

1.4 The Uncanny Valley as a Function of Stimulus and User: The Mixed Effects Approach

To analyze the general phenomenon of the uncanny valley, the common procedure in previous research was to compute the mean of the relevant variables (e.g., eeriness, subjective human likeness) for each stimulus, averaging across all participants (e.g., Burleigh et.al, 2013; MacDorman et al, 2009; MacDorman & Ishiguro, 2006). This procedure corresponds with what Judd and colleagues (2012) call a *by-stimulus analysis*. This procedure treats stimuli as random but ignores random effects due to participants (i.e., systematic variation between persons). As Judd et al. (2012) demonstrated, a by-stimulus analysis will yield positively biased significance tests (alpha inflation) for the effect of an independent variable if in fact participants are random (i.e., differ systematically).

To analyze individual differences in uncanny valley sensitivity, the only existing study directly related to our research (MacDorman & Entezari, 2015) computed a mean on eeriness for each participant, averaging across all stimuli. In the next step, eeriness was predicted by an individual difference construct of interest (e.g., neuroticism). This corresponds to a *by-participant* analysis, which treats participants as random but ignores random effects due to stimuli. If systematic variation between stimuli exists, a by-participant analysis will also yield biased significance tests (Judd et al., 2012).

To avoid collapsing across either stimuli or participants, multilevel modelling (e.g., Raudenbush & Bryk, 2002; Snijders & Bosker, 2012) can be applied to repeated-measures data (i.e., judgments nested in individuals). When a study uses a small number of different (fixed) stimuli, systematic variation between stimuli can be accounted for by including indicator variables for the different stimuli on the judgments level. When a study uses a larger set of stimuli, and these stimuli are sampled from some population of stimuli to which the researchers would like to make inferences, participants and stimuli are both random factors. If all participants are exposed to the same set of stimuli, the two random factors are fully crossed with each other. To account for systematic variation between stimuli in these repeated-measures designs, multilevel models with crossed random effects (e.g., Raudenbush & Bryk, 2002; Snijders & Bosker, 2012) can be applied to the data. Models with crossed random effects are typically called *mixed effects models* (e.g., West, Welch, & Gałecki, 2015). In recent years, different authors (Baayen, Davidson, & Bates, 2008; Judd et al., 2012) have called for an increase in the implementation of these models in experimental and social psychology. As Judd et al. (2012) argued, besides providing unbiased significance tests, "a pronounced benefit [of using a mixed models approach] is that one can obtain estimates of the various variance components, and these may lead in turn to new insights about factors that might be responsible for unexplained variance in data, either associated with stimuli or participants" (p. 65). In our view, there are at least three reasons why research on the uncanny valley phenomenon in particular might profit from applying a mixed effects approach: First, the relationship between human likeness and eeriness can be scrutinized without having to collapse the ratings across participants, and individual differences in uncanny valley sensitivity can be scrutinized without having to collapse the ratings across stimuli. Second, mixed effects models allow researchers to test for not only whether individuals differ in the average level of eeriness that is experienced but also whether individuals differ in the functional form of the relationship between human likeness and eeriness (e.g., in the size of the "drop" in eeriness in the middle of the human likeness continuum). Third, person-level variables can be entered simultaneously into the models to predict individual differences in the uncanny valley experience.

We conducted two experiments that followed the mixed effects rationale. This allowed us to examine whether the relationship between robots' human likeness and respondents' feelings of eeriness followed the pattern suggested by the uncanny valley hypothesis. In Experiment 1, we analyzed eeriness for three human likeness conditions (robot vs. android vs. human) using a small set of different stimuli for each category. We applied a fixed approach to model (potential) systematic differences between stimulus sets in a multivariate multilevel model (with judgments on Level 1 and participants on Level 2). In Experiment 2, we used a larger number of stimuli to represent the human likeness continuum between robots and humans, which allowed us to test for a nonlinear relationship between human likeness and eeriness by applying a polynomial mixed effects model. In both experiments, our focus was on individual differences in the uncanny valley experience. We tested the hypothesis that individuals' need for structure, particularly in the response to lack of structure (RLS) subdimension, would moderate the relationship between human likeness and eeriness responses. Given that RLS, but not DFS, was related to extraversion and neuroticism in previous research (Neuberg & Newsom, 1993), we additionally assessed the two personality traits in Experiment 1 to examine whether the PNS subdimensions' discriminant validity could be confirmed. Moreover, we wanted to test whether the proposed prediction of individual differences in the uncanny valley experience would hold when extraversion and neuroticism (as general susceptibility to positive and negative affectivity) are controlled for (cf. Cavazos et al., 2012).

2. Experiment 1

2.1 Method

2.1.1 Participants.

A total of 331 participants were recruited via social media networks and several university mailing lists in Germany to participate in an online survey on the evaluation of robots. Fiftyseven participants terminated their participation during the task. To identify careless responding (Meade & Craig, 2012), we analyzed completion time and answers to an instructed-response item (“Please respond with totally disagree for this item”). The average completion time was 12.74 min, and 95% of persons spent less than 24.32 min on the survey. The data of 14 participants who took longer than this threshold were excluded from the analyses. At the lower end of the completion time distribution, we excluded the data of the 2.5% fastest participants who completed the survey in less than 4.72 min (n = 7). The data of 22 participants who had not answered the instructed-response item correctly were excluded from the analyses. In addition, we excluded a participant’s data if the participant did not rate the three stimuli on at least two (out of three) eeriness items (n = 4) or if the participant did not complete the PNS, extraversion, and neuroticism items (n = 1). Therefore, the final sample consisted of 226 participants (117 students; 156 women; age range = 15–79, M = 32.00, SD = 13.95).

2.1.2 Stimuli.

Our stimulus pool consisted of three images of robots, three images of humans, and three morphed images (see Figure 2). The depicted robots were selected by their humanoid appearance to make the morphing process easier and to prevent morphing artifacts (i.e., irregularities in the images from the morphing process). Only the faces of the robots and the humans (i.e., not their bodies) were used in the morphing process to make the image composition between human and robot images as similar as possible. The robot images showed the robots called Aila (DFKI Robotics Innovation Center), Bandit (USC Interaction Lab and BlueSky Robotics), or Roboy (AI Lab, University of Zurich), taken from the IEEE robot collection (Institute of Electrical and Electronic Engineers, 2012). Two images of humans were selected from the Karolinska Directed Emotional Faces database (Lundqvist, Flykt, & Öhman, 1998; image IDs AM10NES, BM09NES), and a third image was retrieved via a Google Image search. The images showed the front view of a human face with a direct gaze and a neutral expression. Human and robot images were paired according to facial and feature similarities (e.g., hair, head form, and gender) to simplify the morphing process. After the morphing process, the images were changed to black-and-white, and the contrast was reduced to minimize the morphing artifacts. The image pairs were used as parent faces (i.e., continua endpoints) in the morphing process to produce three stimulus sets. The morphed image we chose for the android category was the image that represented an appearance that was approximately 50% human and 50% robot.

2.1.3 Procedure.

Our experiment was conducted online. First, participants completed the personality measures. Next, we presented the experimental stimuli along with the eeriness and human likeness items (see 2.1.4). Each participant rated three out of the nine images. In counterbalancing the human likeness condition (robot, android, human) and stimulus set, we made sure that each participant rated exactly one robot, one android, and one human and that each image was taken from a different stimulus set. With this procedure, we aimed to reduce potential carry-over effects that occur when individuals rate several images of the same stimulus set. All potential combinations were represented in six different stimulus series, and participants were randomly assigned to one of the series. The presentation order of the three stimuli was determined randomly for each participant. A total of 678 (226 participants x 3 stimuli) eeriness judgments were included in the present analysis. After the stimuli and dependent measures were administered, questions about demographics and media use (unrelated to the current research) followed. Participants were subsequently thanked and debriefed.

2.1.4 Measures.

Eeriness. Eeriness was measured with three items on a 7-point semantic differential scale. The items and format were adapted from Ho and MacDorman's (2010) eeriness-eerie index (1 = *beruhigend* [*reassuring*] to 7 = *furchterregend* [*eerie*]; 1 = *behaglich* [*bland*] to 7 = *unheimlich* [*uncanny*]; 1 = *gruselig* [*creepy*] to 7 = *nicht gruselig* [*not creepy*] [recoded]). The three items were averaged to form an indicator of eeriness. To estimate the reliability of the three-item eeriness measure, we calculated alpha at the within-persons level (level of judgments) and at the between-persons level. We calculated a within-persons-level alpha in Mplus (Muthén & Muthén, 2012) with the method presented by Geldhof, Preacher, and Zyphur (2014); it was .92. On the between-persons level, we calculated Cronbach's alpha separately for each of the nine stimuli[1], which ranged from .70 to .89 (*Mdn* = .84). To make the meaning of the eeriness indicator compatible with the y-axis representing "comfort" in Mori's (1970) conceptualization of the uncanny valley (where low values indicated a negative response and high values indicated a positive response to the stimulus), we reversed the eeriness indicator by multiplying its values by −1. Thus, the transformed eeriness indicator had a possible range from −7 = *very eerie* to −1 = *not at all eerie*. The negative value range offered the advantage that larger absolute scores represented more eeriness (which is intuitively meaningful), but at the same time, a graphical depiction of eeriness on the y-axis was in line with the conceptualization of an uncanny valley (and not an "uncanny mountain").

Subjective human likeness. Participants rated the human likeness of each stimulus with four items on a 7-point semantic differential scale. The items and format were adapted from Ho and MacDorman's (2010) human likeness scale (1 = *künstlich* [*artificial*] to 7 = *natürlich* [*natural*]; 1 = *synthetisch* [*synthetic*] to 7/*echt* [*real*]; 1 = *unbelebt* [*inanimate*] to 7 = *lebendig* [*living*]; 1 = *unmenschlich* [*non-humanlike*] to 7 = *menschlich* [*humanlike*]). The four items were averaged to form an indicator of subjective human likeness. The within-persons-level alpha (Geldhof et al., 2014) was .94. On the between-persons level, we calculated Cronbach's alpha separately for each of the nine stimuli[1], which ranged from .56 to .87 (*Mdn* = .84).

Personal need for structure (PNS). Participants completed the 11-item Personal Need for Structure scale (Neuberg & Newsom, 1993; German version by Machunsky & Meiser, 2006). The subscale Desire for Structure (DFS) consisted of four items (e.g., "I enjoy having a clear and structured way of life"), and the subscale Response to Lack of Structure (RLS) consisted of seven items (e.g., "I don't like situations that are uncertain"). The response format was a 6-point Likert scale (1 = *strongly disagree* to 6 = *strongly agree*). Cronbach's alpha was .73 for DFS and .74 for RLS. For the PNS total scale, Cronbach's alpha was .80.

Extraversion and Neuroticism. Extraversion and Neuroticism were assessed as control variables with the Big Five Inventory (John & Srivastava, 1999; German version by Lang, Lüdtke, & Asendorpf, 2001). The scales consisted of eight (Extraversion) and seven (Neuroticism) items and were rated on a 6-point Likert scale (1 = *strongly disagree* to 6 = *strongly agree*). Cronbach's alpha was .85 for Extraversion and .79 for Neuroticism.

2.1.5 Methods of data analysis.

Our data set included a small number of judgments (three eeriness ratings) per participant—one rating for each of the three human likeness conditions (robot, android, human). We applied a multivariate multilevel modelling approach with judgments on Level 1 and participants on Level 2. A "standard" multivariate multilevel model for a small number of multiple measures (e.g., Snijders & Bosker, 2012) contains as many indicator variables as there are repeated measures. We used this "standard" multivariate model with three dummy variables (one for each judgment) as a baseline model:

Model 0:

Level 1 (judgments):
$$Eerie_{ti} = \beta_{1i} \times Dummy.Robot_{ti} + \beta_{2i} \times Dummy.Android_{ti} + \beta_{3i} \times Dummy.Human_{ti} \quad (0.1)$$

Level 2 (participants):
$$\beta_{1i} = \gamma_{10} + u_{1i} \quad (0.2)$$
$$\beta_{2i} = \gamma_{20} + u_{2i} \quad (0.3)$$
$$\beta_{3i} = \gamma_{30} + u_{3i} \quad (0.4)$$

On Level 1, there is no intercept term so that the varying regression coefficients for the dummy variables (β_{1i}, β_{2i}, and β_{3i}) represent participants' ratings of the three stimuli. Note that the Level-1 equation does not contain a residual term. The reason for this is that participants gave only one rating per type of stimulus. That is, this level serves only as a tool for representing three measures per participant—the dummy variables indicate whether the data line refers to the rating of the android, the robot, or the human stimulus. On Level 2, the fixed effects γ_{10}, γ_{20}, and γ_{30} (eeriness means across participants) are estimated. The variances of the residuals (u_{1i}, u_{2i}, and u_{3i}) represent individual differences in eeriness ratings of the robot, android, and human stimulus, respectively. To test whether an uncanny valley effect emerged, the model can be slightly changed to include an intercept term and two dummy variables which contrast the eeriness mean for the android stimuli with the eeriness means for the robot and human stimuli, respectively:

Model 1:

Level 1 (judgments):
$$Eerie_{ti} = \beta_{0i} + \beta_{1i} \times Dummy.Robot_{ti} + \beta_{2i} \times Dummy.Human_{ti} \quad (1.1)$$

Level 2 (participants):
$$\beta_{0i} = \gamma_{00} + u_{0i} \quad (1.2)$$
$$\beta_{1i} = \gamma_{10} + u_{1i} \quad (1.3)$$
$$\beta_{2i} = \gamma_{20} + u_{2i} \quad (1.4)$$

In this model, the varying intercept term β_{0i} represents participant i's eeriness rating of the android stimulus (reference category), the varying slope term β_{1i} represents the difference between participant i's eeriness rating of the robot stimulus and participant i's eeriness rating of the android stimulus, and the varying slope term β_{2i} represents the difference between participant

i's eeriness rating of the human stimulus and participant i's eeriness rating of the android stimulus. On Level 2, the fixed effects γ_{00} (mean eeriness rating of an android stimulus across participants), γ_{10} (mean difference in eeriness ratings of a robot vs. an android stimulus across participants), and γ_{20} (mean difference in eeriness ratings of a human vs. an android stimulus across participants) are estimated. If we substitute the Level-2 equations into the Level-1 equation, we obtain the following mixed model equation:

$$Eerie_{ti} = \gamma_{00} + u_{0i} + (\gamma_{10} + u_{1i}) \times Dummy.Robot_{ti} + (\gamma_{20} + u_{2i}) \times Dummy.Human_{ti} \tag{2}$$

Due to the small number of selected stimulus sets in this study (three, see Figure 2), we used a fixed effects approach (instead of a random effects approach) to account for possible systematic differences between stimulus sets. To leave the meaning of the intercept, robot slope, and human slope coefficients unchanged (i.e., as described above), we added unweighted effect-coded indicator variables for the stimulus sets (and their interaction terms with the robot and human dummy variables) as Level-1 predictors. This yielded the following mixed model equation:

Model 2:

$$\begin{aligned} Eerie_{ti} = {} & \gamma_{00} + u_{0i} + (\gamma_{10} + u_{1i}) \times Dummy.Robot_{ti} + (\gamma_{20} + u_{2i}) \times Dummy.Human_{ti} \\ & + \gamma_{30} \times Effect.Set2_{ti} + \gamma_{40} \times Effect.Set3_{ti} \\ & + \gamma_{50} \times Dummy.Robot_{ti} \times Effect.Set2_{ti} + \gamma_{60} \times Dummy.Robot_{ti} \times Effect.Set3_{ti} \\ & + \gamma_{70} \times Dummy.Human_{ti} \times Effect.Set2_{ti} + \gamma_{80} \times Dummy.Human_{ti} \times Effect.Set3_{ti} \end{aligned} \tag{3}$$

To test whether the two PNS subscales, RLS and DFS, moderated the uncanny valley experience, we included both PNS subscales (centered at the grand mean) as Level-2 predictors of the intercept term, the robot slope term, and the human slope term. The mixed model equation (with an abbreviation for fixed effects due to stimulus sets) is:

Model 3:

$$\begin{aligned} Eerie_{ti} = {} & \gamma_{00} + u_{0i} + (\gamma_{10} + u_{1i}) \times Dummy.Robot_{ti} + (\gamma_{20} + u_{2i}) \times Dummy.Human_{ti} \\ & [\text{+ fixed effects for stimulus sets}] + \gamma_{01} \times RLS_i + \gamma_{02} \times DFS_i \\ & + \gamma_{11} \times Dummy.Robot_{ti} \times RLS_i + \gamma_{12} \times Dummy.Robot_{ti} \times DFS_i \\ & + \gamma_{21} \times Dummy.Human_{ti} \times RLS_i + \gamma_{22} \times Dummy.Human_{ti} \times DFS_i \end{aligned} \tag{4}$$

where γ_{01} and γ_{02} represent the main effects of the PNS subscales (i.e., the expected increase in eeriness evoked by an android stimulus for a one-unit increase in the PNS subscale, holding the other PNS subscale constant). The fixed effects γ_{11} and γ_{21} represent moderator effects of the RLS subscale (i.e., the expected change in the eeriness difference score [for the robot-android or human-android comparison, respectively] for a one-unit increase in RLS, holding DFS constant). The fixed effects γ_{12} and γ_{22} represent the corresponding moderator effects of the DFS subscale. To test whether the results would hold when Extraversion and Neuroticism were controlled for, we additionally added (grand-mean centered) Extraversion and Neuroticism to Equation 4 as person-level predictors of the varying intercepts and the varying slope coefficients.

We analyzed the data with Mplus (Version 7; Muthén & Muthén, 2012). Mplus allowed us to set the Level-1 variance to zero, which is necessary when specifying a multivariate multilevel model (see e.g., Snijders & Bosker, 2012).

2.2 Results

2.2.1 Manipulation check and descriptive analyses.

Descriptive statistics for the variables can be found in Table 1. As a manipulation check, we analyzed the differences in the subjective human likeness ratings of the robot, android, and human stimuli by using a multivariate multilevel model (cf. Equation 3). That is, dummy variables for the human likeness condition and stimulus set were entered as predictors. The only difference from the model described in Equation 3 was that subjective human likeness (instead of eeriness) was the dependent variable. The results revealed that, as expected, the human stimuli received the highest ratings (5.91), followed by the android stimuli (2.74) and the robot stimuli (1.73). Both the android–robot comparison ($z = -12.22$, $p < .001$) and the android–human comparison ($z = 34.82$, $p < .001$) were significant. The mean human likeness rating of the android stimulus from Stimulus Set 3 (see Figure 2) was higher (by 0.37 points) than the human likeness ratings of the android stimuli on average ($z = 3.39$, $p < .01$). None of the interactions between human likeness condition and stimulus sets were significant (all $|z|$s < 1.86, $p > .06$).

To provide a descriptive overview of the central tendency and the variability in eeriness ratings, Figure 3a depicts the mean eeriness ratings across participants for the three stimulus sets, and Figure 3b depicts boxplots for participants' eeriness ratings as a function of human likeness. On a descriptive level, the android stimuli evoked the highest eeriness. As can be seen in Figure 3a, stimulus sets differed somewhat in the eeriness they evoked. As can be seen in Figure 3b, there were relatively large individual differences in eeriness ratings within human likeness conditions. Estimates of the mean eeriness ratings for each human likeness condition and for the degree of individual differences in these ratings (i.e., *SD*s) were obtained from a baseline multivariate multilevel model. These estimates are depicted in Table 2 (Model 0).

Correlations between the personality variables can be found in Table 3. The PNS subscales were highly positively correlated ($r = .48$). However, only the RLS subscale, but not the DFS subscale, was significantly related to Extraversion and Neuroticism (inverse correlations of moderate size). This suggests that the scales overlap but are not redundant. This is in line with the results of Cavazos et al. (2012), who found that RLS (but not DFS) was significantly positively related to negative trait affectivity and significantly negatively related to positive trait affectivity.

2.2.2 Effect of manipulated human likeness on eeriness.

The results of the multivariate multilevel model that allowed us to test whether the mean eeriness levels for the android versus the robot stimuli and the android versus the human stimuli differed significantly are summarized in Table 2 (Model 1). The eeriness evoked by an android stimulus (−5.51 [γ_{00}]) was significantly stronger (larger negative value) than the eeriness evoked by a robot stimulus (−5.51 [γ_{00}] + 1.51 [γ_{10}] = −4.00) and the eeriness evoked by a human stimulus (−5.51 [γ_{00}] + 2.68 [γ_{20}] = −2.83). Model 2 additionally included effect-coded indicator variables for the stimulus sets to test for differences between the stimulus sets: An android stimulus from Set 2 (see Figure 2) was rated as more eerie (by 0.29 points [γ_{30}]), and an android stimulus from Set 3 (see Figure 2) was rated as less eerie (by 0.30 points [γ_{40}]) compared with the grand mean of eeriness for an android stimulus. Moreover, the android–robot difference was larger for Stimulus Sets 2 and 3 than on average (i.e., for Set 2, the android–robot difference was 1.46 [γ_{10}] + 0.53 [γ_{50}] = 1.99; for Set 3, the android–robot difference was 1.46 [γ_{10}] + 0.40 [γ_{60}] = 1.86), and the android–human difference was smaller for Stimulus Set 3 than on average (i.e., for Set 3, the android–human difference was 2.65 [γ_{20}] − 0.86 [γ_{80}] = 1.79).

2.2.3 Moderator effects of Personal Need for Structure subscales.

Table 2 (Model 3) shows the results for the multivariate multilevel model for predicting eeriness with the person-level PNS subscales. The PNS scales did not show main effects on the eeriness ratings of the android stimuli (nonsignificant γ_{01} and γ_{02}) or moderator effects (nonsignificant γ_{11}, γ_{12}, γ_{21}, and γ_{22}). When only one of the PNS subscales was added as a Level-2 predictor at a time (i.e., either RLS or DFS), the results remained the same. That is, individual differences in the degree to which android stimuli evoked greater eeriness than robot (or human) stimuli could not be explained by individual differences in PNS. When we added Extraversion and Neuroticism as person-level predictor variables to the model, the personality variables did not show main or moderator effects either, and the regression coefficients for the PNS scales did not change (i.e., they remained nonsignificant). When the PNS total scale (instead of the subscales) was analyzed as a person-level predictor variable, a similar picture emerged (there were no main or moderator effects of the PNS total scale).

2.3 Discussion

Adopting a standard procedure in research on the uncanny valley, we asked participants to report their eeriness in response to images that varied in human likeness—that is, robots, humans, and morphed android images. As expected, we found that morphed android images elicited greater eeriness than either robots or humans. This result contributes to the contested question of whether or not the uncanny valley exists. Based on elaborate statistical analyses, our findings (along with Mathur & Reichlich, 2016) support the uncanny valley hypothesis. The effects of the android-human comparison were larger than the effects of the android-robot comparison. This did not come as a surprise given that the selected robots already exhibited humanlike characteristics; thus, the robot's human likeness might have already surpassed the peak of highest affinity. We expected that individual differences in Response to Lack of Structure (RLS) would predict the extent to which the android stimuli would evoke greater eeriness than the robot or human stimuli. Our analyses did not yield an overall effect of RLS on the eeriness ratings, and we did not find the expected moderator effect. Moreover, no influence of recipients' Extraversion or Neuroticism was found in additional exploratory analyses. A limitation of Experiment 1, however, was that the human likeness continuum was represented by three different levels only (robot vs. human vs. one morphed android) with the help of three different stimulus sets. By realizing only three levels of the (theoretically continuous) human likeness dimension, it was not possible to examine the nonlinear effect of more subtle changes in human likeness on eeriness. Thus, we conducted a second experiment with a more fine-grained manipulation of human likeness. Experiment 2 allowed us to include linear, quadratic, and cubic experimental human likeness terms in the analyses as well as terms indicating moderation by the PNS subdimensions.

3. Experiment 2

3.1 Method

3.1.1 Participants.

A total of 547 participants were recruited via social media networks and several university mailing lists in Germany to participate in an online survey on the "evaluation of robots and humans." One hundred ninety-eight participants terminated their participation during the task. That is, they rated only a portion of the stimuli and, more important, they did not complete the PNS measure, which had been placed at the end of the online survey. In addition, to identify careless responding (Meade & Craig, 2012), we analyzed answers to a "Use me" item ("Should we include your data in our analyses?") and completion time. Five participants stated that their data should not be used for data analysis. Average completion time was 19.48 min, and 95% of the participants spent less than 35.7 min on the survey. Data from 18 participants who took longer than this threshold were excluded from the analyses. At the lower end of the completion time distribution, we excluded the data of the 3% fastest participants who completed the survey in less than 6 min (n = 10). Therefore, the final sample consisted of 316 participants (280 students; 194 women; age range = 18−59, M = 24.09, SD = 5.15).

3.1.2 Stimuli.

The stimulus pool consisted of eight stimulus sets with images of robots, humans, and morphed images (see Figure 4, which depicts seven of the eight stimulus sets). Again, robots with a humanoid appearance were chosen to facilitate the morphing process and to minimize the likelihood of morphing artifacts (i.e., irregularities in the images from the morphing process).

As in Experiment 1, the faces of the robots and humans (i.e., not the rest of the body) were used in the morphing process. Three of the selected robot images had already been used in Experiment 1 (Aila, Bandit, Roboy). The five new robot images showed robots named Flobi (University of Bielefeld, Germany), iCub (RoboCub Consortium, Bar Ilan University, Israel), RoboThespain (Engineered Arts Ltd, United Kingdom), and Telenoid (Ishiguro Lab at Osaka University, Japan), taken from the IEEE robot collection (Institute of Electrical and Electronic Engineers, 2012), as well as Roman (Robotics Research Lab, University of Kaiserslautern, Germany). The six male, human images were selected from the Karolinska Directed Emotional Faces database (Lundqvist et al, 1998; image IDs AM10NES, AM18NES, BM09NES, AM30NES, BM25NES, AM31NES), and the two female images were extracted from a Google-Image search, each with a front view of the face with direct gaze and neutral expression. Human and robot images were again paired according to facial and feature similarities (e.g., head form, hair, gender) to simplify the morphing process. The image pairs were then used as parent faces (i.e., stimulus set endpoints) in the morphing process. Each stimulus set consisted of seven levels and was created by increasing the physical human-like appearance and decreasing the physical robot appearance in a stepwise fashion in increments of 16.7% (i.e., Level 1: 0% human/100% robot, Level 2: 16.7% human/83.3% robot ... to Level 7: 100% human/0% robot). For all images, contrast was reduced to minimize artifacts, and the colors were changed to black-and-white.

3.1.3 Procedure.

The study was conducted online and began with an introduction and three example images (one image of each end of the robot-human morphing continuum and the image in the middle of the continuum) so that all participants had the same anchor points for judging human likeness. On the next page, an additional example image was presented along with the human likeness and eeriness items to familiarize participants with the item format. The second part was the actual survey. For each step of the morphing continuum, each participant was presented with four different, randomly selected images. That is, each participant rated 28 images (4 images x 7 steps of the morphing continuum) on human likeness and eeriness. The 28 images were selected randomly out of the 56 existing images (8 different stimulus sets x 7 steps) and were presented in a random order for each participant. In the last part of the survey, participants were requested to complete the PNS measure and answer demographic questions. On the last page, participants were informed about the purpose of the survey and had the opportunity to withdraw their data from data analysis.

3.1.4 Measures.

Eeriness. Eeriness was measured with two items (unheimlich [uncanny] and gruselig [eerie]) taken from Ho and MacDorman's (2010) eeriness-eerie index. For each image, participants rated the degree of eeriness the stimulus evoked on a 7-point intensity scale (e.g., 1 = *not at all eerie* to 7 = *very eerie*). Given that the number of stimuli to be rated was relatively large in this experiment (compared with Experiment 1), we used only two items to measure eeriness to reduce participant burden. The two items were averaged to form an indicator of eeriness. To estimate reliability, we calculated alpha at the within-persons level (level of judgments) and at the between-persons level in Mplus (Muthén & Muthén, 2012) by applying the method presented by Geldhof et al. (2014) to our data[2]. For the two-item measure of eeriness, alpha was .97 on the within-persons (judgments) level and .98 on the between-persons level. As in Experiment 1, we reversed the eeriness indicator by multiplying its values by −1. Thus, the transformed eeriness indicator had a possible range from −7 = *very eerie* to −1 = *not at all eerie*.

Subjective human likeness. For each image, participants rated the stimulus' degree of human likeness on two items (künstlich–natürlich [artificial–lifelike] and unmenschlich–menschlich [machinelike–humanlike]) using a 7-point bipolar intensity scale (e.g., 1 = *very artificial* to 7 = *very lifelike*). The items were taken from the five-item human likeness scale used by Ho and MacDorman (2010) and Mara and Appel (2015a, 2015b). Again, we used only two items to measure human likeness in order to reduce participant burden. The two items were averaged to form an indicator of human likeness. To estimate reliability, we calculated alpha at the within-persons level (level of judgments) and at the between-persons level (Geldhof et al., 2014). For the two-item measure of human likeness, alpha was .95 on the within-persons (judgments) level and .65 at the between-persons level.

Personal Need for Structure (PNS). We used the same Personal Need for Structure scale that was used in Experiment 1. Cronbach's alpha was .73 for the Desire for Structure subscale, .65 for the Response to Lack of Structure subscale, and .77 for the PNS total scale.

3.1.5 Methods of data analysis.

In the final sample of 316 participants, a total of 59 responses to individual items (0.1% of all responses) were missing. Because this proportion was rather small, we decided to deal with missing data at the item level by using the available items as indicators of a stimulus' human likeness or eeriness. For the PNS scale, the available items were averaged to form the scale score. In total, 8,846 judgments of images were analyzed in the present analyses.

Our data set represents repeated-measures data (judgments on Level 1) with participants (Level 2a) and stimulus sets (Level 2b) as crossed random effects. Given that for each step of the manipulated human likeness continuum, each participant was presented with four randomly selected images, not all participants were exposed to all stimulus sets (although the majority were: 91% of the participants saw images from all eight stimulus sets). Thus, participants and stimulus sets were partially crossed in our design.

To test whether a curvilinear relationship between human likeness and eeriness held in our data, a mixed model specifying a cubic function for human likeness on Level 1 (i.e., an S-shaped function with two bends) was used. The mixed model equation is:

$$Eerie_{tij} = \gamma_{00} + u_{0i} + v_{0j} + (\gamma_{10} + u_{1i} + v_{1j}) \times HL_{tij} + (\gamma_{20} + u_{2i} + v_{2j}) \times HL_{tij}^{2} + (\gamma_{30} + u_{3i} + v_{3j}) \times HL_{tij}^{3} + \varepsilon_{tij} \quad (5)$$

where $Eerie_{tij}$ represents the eeriness ratings t (t = 1, 2, …, 28) for participant i (i = 1, 2, …, 344) and stimulus set j (j = 1, 2, …, 8). γ_{00} is the fixed intercept, u_{0i} is the random effect for the intercept associated with participant i, and v_{0j} is the random effect for the intercept associated with stimulus set j. The fixed effects γ_{10}, γ_{20}, and γ_{30} characterize the average relationship between human likeness and eeriness (across all participants and all stimulus sets). Depending on which of these three fixed effects reaches significance, the relationship between human likeness and eeriness follows a linear, a quadratic, or a cubic function. The highest significant polynomial term defines the form of the relationship (Cohen, Cohen, West, & Aiken, 2003). Random effects for the three human likeness terms were specified across participants (u_{1i}, u_{2i}, and u_{3i} representing between-persons differences in the human likeness–eeriness relationship) and across stimulus sets (v_{1i}, v_{2i}, and v_{3i} representing between-stimuli differences in the human likeness–eeriness relationship), and we tested (via deviance tests) whether these random variance components were significantly larger than zero. Only significant random effects were retained in the model (see e.g., West et al., 2015). The residual term ε_{tij} represents the deviations of eeriness ratings from the expected eeriness score for a given person–stimulus set combination. In the first set of models, we used manipulated human likeness (steps of the morphing continuum) as the Level-1 predictor variable. Manipulated human likeness was centered to facilitate interpretation of the mixed model coefficients (thus, it ranged from −3 = robot stimulus to 3 = human stimulus). In a second set of models, we used subjectively rated human likeness (instead of manipulated human likeness steps) as the Level-1 predictor variable. We centered subjectively rated human likeness at the person mean so that pure within-persons relationships would be estimated, and the interpretation of the regression coefficients would be facilitated (Enders & Tofighi, 2007).

To test whether the two PNS subscales, Response to Lack of Structure (RLS) and Desire for Structure (DFS), moderated the uncanny valley experience, we included the two person-level variables (centered at the grand mean) as predictors of the intercept and of the linear, quadratic, and cubic manipulated human likeness terms, respectively. The mixed model equation is:

$$\begin{aligned} Eerie_{tij} &= \gamma_{00} + u_{0i} + v_{0j} + (\gamma_{10} + u_{1i} + v_{1j}) \times HL_{tij} + (\gamma_{20} + u_{2i} + v_{2j}) \times HL_{tij}^2 + (\gamma_{30} + u_{3i} + v_{3j}) \times HL_{tij}^3 \\ &+ \gamma_{01} \times RLS_i + \gamma_{11} \times RLS_i \times HL_{tij} + \gamma_{21} \times RLS_i \times HL_{tij}^2 + \gamma_{31} \times RLS_i \times HL_{tij}^3 \\ &+ \gamma_{02} \times DFS_i + \gamma_{12} \times DFS_i \times HL_{tij} + \gamma_{22} \times DFS_i \times HL_{tij}^2 + \gamma_{32} \times DFS_i \times HL_{tij}^3 + \varepsilon_{tij} \end{aligned} \tag{6}$$

where the fixed effects γ_{01} and γ_{02} represent the main effects of the PNS subscales (i.e., the change in eeriness of a stimulus that is at the midpoint of the human likeness continuum for a one-unit increase in one of the subscales, holding the other subscale constant). The fixed effects γ_{11}, γ_{21}, and γ_{31} represent the moderator terms for the first subscale (RLS), and the fixed effects γ_{12}, γ_{22}, and γ_{32} represent the moderator terms for the second subscale (DFS). In a separate model, subjective (instead of manipulated) human likeness was entered as a Level-1 predictor variable (again, centered at the person mean). In this model, the fixed effects have the same general meaning as in the model that included manipulated human likeness as the Level-1 predictor variable (the only difference is that the intercept represents the expected eeriness of a stimulus at the average level of subjective human likeness). Additionally, we analyzed all models with only one of the PNS subscales included as a predictor variable (instead of including both subscales simultaneously). This was done to scrutinize whether using the total variance of a PNS subscale as a predictor (vs. the unique variance that is not shared with the other subscale) leads to different results.

We analyzed the data with a mixed effects modelling approach in R (R Core Team, 2015) using the lmer() function from the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). Degrees of freedom and *p*-values for the fixed effects were determined by applying the R-package lmerTest (Kuznetsova, Brockhoff, & Christensen, 2015).

3.2 Results

3.2.1 Descriptive analyses.

Descriptive statistics for all variables can be found in Table 1, and the correlations between the PNS scales (subscales, total scale) can be found in Table 3. To provide a descriptive overview of the variability in eeriness ratings among stimulus sets and participants, we generated boxplots for eeriness as a function of manipulated human likeness (i.e., the seven steps of the morphing continuum).

Figure 5a depicts the by-stimulus set analysis and Figure 5b the by-participant analysis. As can be seen in both Figures 5a and 5b, on a descriptive level, the relationship between human likeness and eeriness was curvilinear, with stronger eeriness evoked by the midpoint than by the poles of the human likeness continuum. Figure 5a provides evidence of some amount of variability in eeriness ratings among stimulus sets (with larger variability across stimuli near the robot pole than near the human pole of the human likeness continuum). This suggests that a mixed effects model of these data should include a random effect (random intercept) for stimulus sets. Figure 5b demonstrates that the variability in eeriness ratings among participants was very large, suggesting that a random effect (random intercept) for participants should also be included in a mixed model for these data. Two-level null models (without predictor variables) with crossed random effects for participants and stimulus sets revealed that 19% of the variance in eeriness ratings was due to differences between participants, and 3% of the variance in eeriness ratings was due to differences between stimulus sets.

Next, we analyzed the subjective human likeness ratings as a function of manipulated human likeness (i.e., the seven steps of the morphing continuum). The subjective human likeness ratings increased in a monotonic fashion across the steps of the morphing continuum. The within-persons correlation between manipulated human likeness and subjective human likeness was very large, $r = .85$ ($p < .001$). Two-level null models with crossed random effects for participants and stimulus sets revealed that 3% of the variance in the subjective human likeness ratings was due to differences between participants and 3% of the variance in subjective human likeness ratings was due to differences between stimulus sets. That is, systematic differences between participants were much smaller for the subjective human likeness ratings than for the eeriness ratings.

3.2.2 Test of the functional form of the human likeness–eeriness relationship.

Model comparisons revealed that in a mixed model predicting eeriness on Level 1 (Judgments Level) by *manipulated human likeness*, crossed random intercept effects and crossed random effects for the linear and quadratic terms (but not for the cubic term) of human likeness had to be specified for participants and stimulus sets. As can be seen from the fixed part of Model 1 in Table 4, the cubic human likeness term (γ_{30}) was significant. This means that the average relationship between human likeness and eeriness could be described by a cubic function (i.e., a function with two bends).

In Figures 6a and 6b, the shape of this (average) curvilinear relationship is depicted by a bold curve. Note that the cubic term determines the shape of the function within the range between $-\infty$ and $+\infty$. Only one bend in the curve is visible in the possible range of the (centered) human likeness continuum. This bend represents the expected uncanny valley. The second bend in the cubic function would be visible only if a larger range of positive values on the right were depicted. This might be interpreted as showing that for the humanlike pole of the x-axis, the average eeriness ratings (asymptotically) approached the highest possible score (the "not at all eerie" category). As a measure of effect size, we calculated the proportional reduction in the Level-1 variance when human likeness (linear, quadratic, and cubic) was added to the model (quasi-R^2 for Level 1; see e.g., Snijders & Bosker, 2012). It was .45—that is, compared with the null model, predicting eeriness with a cubic function for human likeness reduced the unexplained variance in eeriness ratings by 45% (which can be considered a large effect). Between-stimulus-set differences (i.e., random effects across stimulus sets) are illustrated in Figure 6a, and between-participant differences (i.e., random effects across participants) are illustrated in Figure 6b.

When *subjective human likeness* was used as the Level-1 predictor variable (see Model 1 in Table 5), the results were very similar to what we found for manipulated human likeness as the Level-1 predictor variable. Due to page restrictions, we decided not to include extra figures for the cubic relationship between subjective human likeness and eeriness.

3.2.3 Moderator effects of Personal Need for Structure subscales.

To test whether the two PNS subscales (DFS and RLS) would predict individual differences in uncanny valley sensitivity, we first entered both PNS subscales as person-level predictors in the model with *manipulated* human likeness as the Level-1 predictor variable. As can be seen in Model 2 in Table 4, the two subscales demonstrated both main effects (i.e., predicting the varying intercept term) and interaction effects (predicting the linear and quadratic terms of manipulated human likeness). It is important to note that these effects had different signs: RLS negatively predicted the intercept term, and DFS positively predicted the intercept term. Inverse regression coefficients were also found for the prediction of the quadratic term of manipulated human likeness.[3] To gain more insight into the form these main and interaction effects took, we plotted the predicted curves for individuals with low (*M* – 1 *SD*), medium (*M*), and high (*M* + 1 *SD*) RLS scores (see Figure 7a) as well as for individuals with low (*M* – 1 *SD*), medium (M), and high ($M + 1$ SD) DFS scores (see Figure 7b). As expected, the uncanny valley was more pronounced (steeper) for individuals high on RLS than for individuals low on RLS. For DFS, the pattern was reversed: The uncanny valley was more pronounced for individuals low on DFS than for individuals high on DFS. When only RLS was included as a person-level predictor (see Table 4, Model 3), the direction of the effects remained the same, but the regression coefficients were somewhat smaller than in the simultaneous analysis. The main effect of RLS and the linear interaction effect remained significant, and additionally, the cubic interaction term reached significance. That is, the form of the moderator effect of RLS on uncanny valley sensitivity remained very similar. When only DFS was included (see Table 4, Model 4), the main and the interaction effect terms were no longer significant.

When the PNS subscales were entered as person-level predictors in the model that included *subjective* (instead of manipulated) human likeness as the Level-1 predictor variable, the overall pattern of results was similar (see Table 5, Model 2): The main effects of the PNS subscales were again significant and opposite in sign. That is, for individuals high on RLS, the stimuli of perceived android quality evoked more eeriness (larger negative predicted scores) than for individuals low on RLS, and the opposite pattern held for individual differences in the DFS facet. However, with the exception of the interaction between DFS and the quadratic term, which was significant, the other interaction effects were somewhat smaller than in the corresponding Model 2 in Table 4 and did not reach the .05 level of significance ($p < .10$). Because the significant main effects and the significant interaction effect were similar to the results found for manipulated human likeness, we refrained from including separate plots for subjective human likeness. When either RLS or DFS were included as a predictor (see Table 5, Models 3 and 4), the direction of the effects again remained the same, and the regression coefficients were smaller than in the simultaneous analysis. For RLS, the main effect and the quadratic interaction effect missed the .05 level of significance ($p < .10$). For DFS, only the quadratic interaction term remained significant.

3.3 Discussion

We found a cubic relationship between robots' human likeness and participants' eeriness ratings. However, only one bend in the estimated curve—representing the uncanny valley—was located in the possible range of the human likeness continuum. The second bend in the estimated curve would be visible only if a larger range of positive values on the right were present. This suggests that for the human endpoint of the x-axis, average eeriness ratings approached minimum eeriness. Our results fit the uncanny valley proper (cf. Kätsyri et al., 2015), but the left hand upward slope of Mori's graph was omitted (see Moore, 2012, and Yamada et al., 2013, for similar findings). Moreover, the point of maximum eeriness was closer to the robot endpoint of the continuum than to the human endpoint. Both findings are likely due to the fact that the examples used as the robot endpoints of the morphing continua were already moderately humanlike. They showed substantial human features and proportions (which facilitated the morphing process and reduced potential artifacts).

The observed main and interaction effects of the RLS subscale point out that the uncanny valley was similar in shape but more pronounced for high RLS participants, and it was similar in shape but less pronounced for low RLS participants. We found evidence for the hypothesized moderator effect of RLS when we used manipulated human likeness (i.e., steps of the morphing continuum) to model the uncanny valley phenomenon. When we used subjective human likeness ratings to model the uncanny valley phenomenon, a similar moderator effect of RLS emerged, but the empirical evidence was weaker. This might be due to the fact that subjective human likeness ratings (although highly correlated with manipulated human likeness) introduce more random variance into the uncanny valley part of the model, which might make it more difficult to detect a moderator effect of RLS on the shape of the human likeness–eeriness link.

It is interesting that the moderator effect of RLS took the same form but was larger when the second PNS subscale (DFS) was controlled. That is, the unique variance of RLS, in particular, contributed to the prediction of individual differences in uncanny valley sensitivity. According to Neuberg and Newsom (1993), a general preference for simple (as opposed to complex) structures represents the shared variance of the RLS and DFS subdimensions, and the tendency to react negatively to the absence of structure represents the unique aspect of RLS. Previous research has revealed that RLS (but not DFS) is associated with negative trait affectivity and sensitivity to the negative emotional qualities of stimuli (Cavazos et al., 2012). Moreover, Elovaino and Kivimäki (1999) have shown that RLS (but not DFS) is associated with higher occupational strain among individuals working in occupations with high complexity. Our findings provide first evidence that being confronted with humanlike robots elicits individual differences in eeriness (i.e., a specific type of negative affective response) which are related to individual differences in RLS. Hence, our finding is in line with previous research on the discriminant validity of the PNS subdimensions. Unexpectedly, we found some evidence for an inverse moderator effect of DFS when RLS was controlled. Given that we had not formulated a moderator hypothesis for the DFS subdimension, we would prefer to treat this as an exploratory finding which warrants further investigation.

Taken together, the findings of Experiment 2 indicate that the valley-like form of the relationship between robots' human likeness and individuals' eeriness held for all participants. We found some evidence that the deepest point in the valley varied with individuals' tendency to respond negatively to a lack of structure.

4. General Discussion

4.1 Contribution

The uncanny valley hypothesis (Mori, 1970 / Mori, MacDorman, & Kageki, 2012), which suggests a nonlinear relationship between human likeness and eeriness, has received substantial attention from scholars, the popular media, as well as the general public, but evidence for its mere existence, let alone for its predictors and the underlying mechanisms, is far from established. In previous research, the reactions towards humanlike robots have been attributed almost exclusively to the robot's design, including its nonverbal behavior (Mara & Appel, 2015a) and the fit between visual appearance and movement (Saygin et al., 2012). From a psychological standpoint, such an approach is insufficient. Individuals differ in their responses to given robotic stimuli, and largely ignoring these differences is a substantial shortcoming of the field. Not only does this practice obfuscate the phenomenon conceptually and empirically, it can lead to biased estimates of the stimulus effects (Judd et al., 2012). On the basis of current theories about the psychological mechanisms underlying the uncanny valley effect, we examined whether the personality trait need for structure (Neuberg & Newsom, 1993) could explain individual differences in eeriness responses. Our work is the first set of experiments to be based on a mixed effects rationale with a focus on nonlinear functions. This allowed us to examine the relationship between human likeness and eeriness without collapsing eeriness ratings across participants and to examine individual differences in uncanny valley sensitivity without having to collapse eeriness ratings across stimuli (as has been done in previous research). It is important to note that the nonlinear mixed effects approach allowed us to model individual differences in the functional form of the relationship between human likeness and eeriness and to test whether individual differences in the need for structure would predict these functional differences in the uncanny valley experience.

The results of our statistical analyses supported the uncanny valley hypothesis. We found that an android elicited stronger feelings of eeriness than a moderately humanlike robot or a human (Experiment 1) and that along the continuum between a moderately humanlike robot and a human being, the relationship between human likeness and eeriness followed a cubic function (Experiment 2). Moreover, the findings of both experiments showed that individuals differed substantially in their eeriness reactions, a finding that underscores the need to acknowledge this previously largely neglected source of variance in future analyses of the uncanny valley phenomenon.

As demonstrated in Experiment 2, individual differences in the functional form of the uncanny valley experience could be predicted by a subdimension of the need for structure: Individuals who tend to react negatively whenever structure is disrupted (Response to Lack of Structure [RLS]; Neuberg & Newsom, 1993) had stronger negative reactions to the ambiguous stimuli. This disposition was related to more intense uncanny feelings and a more pronounced uncanny valley effect. It is important to note that the second subdimension of need for structure, Desire for Structure (DFS), did not show a similar moderator effect in Experiment 2. When RLS was controlled, DFS even predicted less eeriness in response to android stimuli. However, given the unexpectedness of this result, the role of DFS in eeriness experiences should be scrutinized in future research. Nonetheless, the findings of Experiment 2 underscore the importance of distinguishing between the two subcomponents of the need for structure (Cavazos et al., 2012; Neuberg & Newsom, 1993), which might substantially increase the predictive power of this trait.

It needs to be emphasized that the predictive power of RLS in Experiment 2 was more pronounced for the reactions to the android stimuli (at the bottom of the uncanny valley, near the midpoint of the selected morphing continuum) than for the reactions to the more clear-cut robot or human stimuli (at the endpoints of the selected morphing continuum). The observed effects build upon and expand previous results on bivariate relations between personality traits and eeriness ratings (MacDorman & Entezari, 2015). In general, bivariate relations with "uncanny valley sensitivity" might simply reflect a tendency to respond with eeriness to all kinds of stimuli, including nonhumanlike robots and humans. Although MacDorman and Entezari (2015) reported personality correlates of eeriness ratings for one class of stimuli (androids) only, they noted (in a footnote) that for eeriness ratings of nonandroids, only a few relations with personality variables were significant. By analyzing a larger spectrum of the human likeness continuum simultaneously, we were able to show that RLS predicted eeriness responses to a larger degree near the middle of our human likeness continuum than at the endpoints of the continuum, thereby ruling out a simple "eeriness-proneness" explanation.

Although our hypothesis on the association between RLS and uncanny valley sensitivity was confirmed in Experiment 2, we did not find evidence for this hypothesis in Experiment 1. Hence, the moderator effect of RLS in Experiment 2 requires cautious interpretation and awaits replication in future research. We can only speculate about reasons why the two experiments did not yield converging evidence on the role of RLS: First, Experiment 2 used a fine-grained (seven-step) manipulation of human likeness, but Experiment 1 used a rather coarse (three-step) manipulation of human likeness (only one selected morphed stimulus plus the endpoints of the continuum). This might be interpreted to point out the importance of selecting a broad spectrum of stimuli when analyzing individual differences in the uncanny valley experience. Second, in Experiment 2, the relation between RLS and eeriness responses was most pronounced for stimuli that were located somewhat closer to the robot endpoint of the human likeness continuum than to the human endpoint of the continuum (see Figure 7a). However, the morphed stimuli selected for the android category in Experiment 1 were located at the midpoint of the human likeness continuum. That is, android stimuli in Experiment 1 contained a larger proportion of human features than those stimuli in Experiment 2 that evoked the largest personality-consistent individual differences in eeriness responses.

4.2 Limitations and Future Research

Using elaborate statistical methods, we found support for the uncanny valley hypothesis. Our goal had not been to rule out a specific theory on the psychological mechanisms underlying the uncanny valley effect or to compare different theories and decide whether the data supported one or the other. Our research provides tentative evidence for the importance of individual differences in the need for structure in the formation of the uncanny valley experience. Thus, our results are in line with category uncertainty theory and the expectancy violation approach as well as with a terror management account (MacDorman, 2005; Landau et al., 2004), or the meaning violation framework (Mara & Appel, 2015b; Heine, Proulx, & Vohs, 2006)—theories and models with a great deal of overlap (Proulx, Inzlicht, & Harmon-Jones, 2012).

Our decision to examine the need for structure as the individual difference variable was guided by current theories on the psychological origins of the uncanny valley phenomenon. Other personality variables might qualify for theory-guided hypotheses as well (cf. MacDorman & Entezari, 2015). One such personality variable could be the need for affect (Maio & Esses, 2001), which is related to a particular sensitivity to the warm/cold dimension in person perception (Aquino, Haddock, Maio, Wolf, & Alparone, 2016), a dimension that is closely connected to the uncanny valley experience. Regarding individual differences beyond the realm of personality, cultural context has received surprisingly little attention in research on the uncanny valley. Anecdotal evidence suggests that Asian—and especially Japanese—people are more accepting of humanlike robots than persons socialized in other parts of the world. Future research should investigate whether the sensitivity to humanlike robots is indeed influenced by cultural factors.

Like much of the research on the uncanny valley, we used images that represented a morphing continuum from robot to human (cf. Burleigh et al., 2013; Hansen, 2006; MacDorman

& Ishiguru, 2006; MacDorman et al., 2009; Yamada et al., 2013) as our stimulus material. We acknowledge that morphing artifacts can reduce the validity of this method. We reduced morphing artifacts during stimulus generation; thus, we have reason to believe that our effects did not depend on this methodological shortcoming. An alternative approach is the use of images of existing (humanlike) robots as stimuli. Although this approach would have increased the study's external validity, the choice of the robots is crucial to the method and could determine the results: Whereas one recent study based on this approach found evidence for an uncanny-valley-like function between human likeness and eeriness (Mathur & Reichling, 2015), others found no such evidence (Rosenthal-von der Pütten & Krämer, 2014). As humanlike robots become more and more available to researchers, we envisage that more experimental studies will examine real-life human-robot interactions, a setting that has rarely been used in theory-guided experimental research with substantial sample sizes (for an exception, see e.g., Mara & Appel, 2015b).

4.3 Conclusion

Feelings of eeriness in response to robots are traditionally conceived of as a function of the robotic stimulus alone. This perspective is incomplete and potentially erroneous from a statistical perspective (e.g., Judd et al., 2012). Using a nonlinear mixed effects approach, we showed that the uncanny valley is a reliable phenomenon. Differences between individuals were nonetheless substantial. In one of two experiments, we found support for the hypothesis that individuals who react more negatively to a lack of structure (Neuberg & Newsom, 1993) demonstrate higher uncanny valley sensitivity. Replication studies might profit from augmenting the stimulus set to enhance power (Westfall, Kenny, & Judd, 2014). Future research on the uncanny valley phenomenon should account for both person and contextual factors when investigating user responses to (humanlike) robots.

References

Aquino, A., Haddock, G., Maio, G. R., Wolf, L. J., & Alparone, F. R. (2016). The role of affective and cognitive individual differences in social perception. *Personality and Social Psychology Bulletin, 42*, 798–810. doi: 10.1177/0146167216643936

Azarian, B. (2015, October). Avoiding the Uncanny Valley [Journalistic Article]. Retrieved from http://www.slate.com/articles/technology/future_tense/2015/10/research_supports_the_uncanny_valley_theory_of_human_robot_interaction.html

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412. doi:10.1016/j.jml.2007.12.005

Bar-Tal, Y., & Guinote, A. (2002). Who exhibits more stereotypical thinking? The effect of need and ability to achieve cognitive structure on stereotyping. *European Journal of Personality, 16*, 313–331. doi: 10.1002/per.453. doi:10.1002/per.453

Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2007, August). Is the uncanny valley an uncanny cliff? In *Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communications (RO-MAN 2007)* (pp. 368–373). Jeju, Korea: IEEE. doi:10.1109/ROMAN.2007.4415111

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2016). *lme4: Linear Mixed-Effects Models using 'Eigen' and S4* (R package version 1.1–11). Retrieved from https://cran.r-project.org/web/packages/lme4/

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

Cavazos, J. T., Judice-Campbell, N., & Ditzfeld, C. P. (2012). Differing emotional sensitivities in the two factors of personal need for structure. *Journal of Research in Personality*, *46*, 49–54. doi:10.1016/j.jrp.2011.12.005

Cheetham, M., Suter, P., & Jäncke, L. (2011). The human likeness dimension of the “uncanny valley hypothesis”: Behavioral and functional MRI findings. *Frontiers in Human Neuroscience, 5,* article 126. doi:10.3389/fnhum.2011.00126

Clow, K. A., & Esses, V. M. (2005). The development of group stereotypes from descriptions of group members: An individual difference approach. *Group Processes & Intergroup Relations, 8*, 429–445. doi:10.1177/1368430205056469

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. New York: Routledge.

Elovainio, M., & Kivimaki, M. (1999). Personal need for structure and occupational strain: An investigation of structural models and interaction with job complexity. *Personality and Individual Differences*, *26*, 209–222. doi: 10.1016/S0191-8869(98)00038-5

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*, 121–138. doi:10.1037/1082-989X.12.2.121

Eveleth, R. (2013, September 2). Robots: Is the uncanny valley real? [Journalistic Article]. Retrieved from http://www.bbc.com/future/story/20130901-is-the-uncanny-valley-real

Frenkel-Brunswik, E. (1949). Intolerance of ambiguity as an emotional and perceptual personality variable. *Journal of Personality*, *18*, 108–143. doi:10.1111/j.1467-6494.1949.tb01236.x

Gates, B. (2007). A robot in every home. *Scientific American*, *296*, 58–65.

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*, 72–91. doi:10.1037/a0032138

Gocłowska, M. A., Baas, M., Crisp, R. J., & De Dreu, C. K. (2014). Whether social schema violations help or hurt creativity depends on need for structure. *Personality and Social Psychology Bulletin*, *40*, 959–971. doi: 10.1177/0146167214533132

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*, 619. doi: 10.1126/science.1134475

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

Hanson, D. (2005). Expanding the aesthetics possibilities for humanlike robots. In *Proceedings of the IEEE Humanoid Robotics Conference, Special Session on the Uncanny Valley*. Tsukuba, Japan.

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

Heine, S. J., Proulx, T., & Vohs, K. D. (2006). The meaning maintenance model: On the coherence of social motivations. *Personality and Social Psychology Review*, *10*, 88–110. doi:10.1207/s15327957pspr1002_1

Hirai, K., Hirose, M., Haikawa, Y., & Takenaka, T. (1998, May). The development of Honda humanoid robot. In *Proceedings of the international conference on robotics and automation (Vol. 2,* pp. 1321–1326). Leuven, Belgium: IEEE Press. doi:10.1109/ROBOT.1998.677288

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

Hurst, S. (2014, September 15). Family robot Jibo finishes indiegogo run; raises over $2M & becomes most successful technology campaign on the platform. *Crowdfund Insider.* Retrieved from http://www.crowdfundinsider.com/

Institute of Electrical and Electronic Engineers (IEEE). (2012). Robots (Version 1.6) [Mobile application Software]. Retrieved from http://itunes.apple.com

International Federation of Robotics (2015). *World robotics survey: Service robots are conquering the world.* Retrieved from http://www.worldrobotics.org/index.php?id=home&newsid=285

Jentsch, E. (1906). Zur Psychologie des Unheimlichen. *Psychiatrisch-neurologische Wochenschrift, 8,* 195–198, 203–205. [English version: Jentsch, E. (1997). On the psychology of the uncanny. *Angelaki: Journal of the Theoretical Humanities, 2,* 7–16.] doi:10.1080/09697259708571910

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York, NY: Guilford.

Jonette (2016, May). Uncanny valley gap: Humanoid robots, or cute robots [Blog Post]. Retrieved from http://socialnewsdaily.com/62789/uncanny-valley-gap-humanoid-robots-or-cute-robots/

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: a new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*, 54–69. doi: 10.1037/a0028347

Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology*, *6*. doi: 10.3389/fpsyg.2015.00390

Kruglanski, A. W., Webster, D. M., & Klem, A. (1993). Motivated resistance and openness to persuasion in the presence or absence of prior information. *Journal of Personality and Social Psychology, 65*, 861–876. doi:10.1037/0022-3514.65.5.861

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). *lmerTest: Tests in Linear Mixed Effects Models* (R-package version 2.0–30). Retrieved from https://cran.r-project.org/web/packages/lmerTest/

Landau, M. J., Johns, M., Greenberg, J., Pyszczynski, T., Martens, A., Goldenberg, J. L., et al. (2004). A function of form: Terror management and structuring the social world. *Journal of Personality and Social Psychology, 87*, 190–210. doi:10.1037/0022-3514.87.2.190

Lang, F. R., Lüdtke, O., & Asendorpf, J. (2001). Testgüte und psychometrische Äquivalenz der deutschen Version des Big Five Inventory (BFI) bei jungen, mittleren und alten Erwachsenen. [Psychometric properties and equivalence of a German version of the Big Five Inventory (BFI) for young, middle aged, and older adults]. *Diagnostica, 47*, 111–121. doi:10.1026//0012-1924.47.3.111

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces (KDEF).* CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9.

MacDorman, K. F. (2006, July). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. In *Proceedings of the*

ICCS/CogSci-2006 Long Symposium: Toward social mechanisms of android science (pp. 26–29).

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

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

MacDorman, K. F., Green, R. D., Ho, C. C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior, 25,* 695–710. doi:10.1016/j.chb.2008.12.026

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies, 7*, 297–337. doi:10.1075/is.7.3.03mac

Machunsky, M., & Meiser, T. (2006). Personal Need for Structure als differenzialpsychologisches Konstrukt in der Sozialpsychologie: Psychometrische Analyse und Validierung einer deutschsprachigen PNS-Skala [Personal need for Structure as an individual differences construct in social psychology: Psychometric analysis and validation of a German version of the PNS scale]. *Zeitschrift für Sozialpsychologie*, *37*, 87–97. doi:10.1024/0044-3514.37.2.87

Maio, G. R., & Esses, V. M. (2001). The Need for Affect: Individual differences in the motivation to approach or avoid emotions. *Journal of Personality, 69*, 583–615. doi: 10.1111/1467-6494.694156

Mara, M. (2014). *Why humanlike robots give us the creeps: An integrative review of the “Uncanny Valley” research.* Unpublished Manuscript. University of Koblenz-Landau, Germany.

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

Mara, M., & Appel, M. (2015b). Science fiction reduces the eeriness of android robots: A field experiment. *Computers in Human Behavior, 48*, 156–162. doi:10.1016/j.chb.2015.01.007

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

Matsui, D., Minato, T., MacDorman, K. F., & Ishiguro, H. (2005, August). Generating natural motion in an android by mapping human motion. In *Proceedings of the International Conference on Intelligent Robots and Systems (IROS 2005)* (pp. 3301–3308). IEEE. doi:10.5772/4812

McGregor, I., Haji, R., & Kang, S. J. (2008). Can ingroup affirmation relieve outgroup derogation? *Journal of Experimental Social Psychology*, *44*(5), 1395–1401. doi:10.1016/j.jesp.2008.06.001

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*, 437–455. doi: 10.1037/a002808

Meth, D. (2014, August) Welcome to Uncanny Valley …where robots make you want to throw up [Blog Post]. Retrieved from https://www.buzzfeed.com/danmeth/welcome-to-uncanny-valley?utm_term=.fp1BgGEzmA#.yjWBDmOkq9

Mitchell, W. J., Szerszen Sr, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception, 2*, 10. doi:10.1068/i0415

Moore, R. K. (2012). A Bayesian explanation of the ‘Uncanny Valley’ effect and related psychological phenomena. *Scientific Reports*, 2. doi:10.1038/srep00864

Mori, M. (1970). Bukimi no tani (the uncanny valley). *Energy*, *7*, 33–35. Translated in: Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *Robotics & Automation Magazine, IEEE*, *19*(2), 98–100. doi: 10.1109/mra.2012.2192811

Moskowitz, G. B. (1993). Individual differences in social categorization: The influence of personal need for structure on spontaneous trait inferences. *Journal of Personality and Social Psychology, 65*, 132–142. doi:10.1037/0022-3514.65.1.132

Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Neuberg, S. L., & Newsom, J. T. (1993). Personal need for structure: Individual differences in the desire for simple structure. *Journal of Personality and Social Psychology, 65*, 113–131. doi:10.1016/j.sbspro.2014.12.458

Piwek, L., McKay, L. S., & Pollick, F. E. (2014). Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition*, *130*, 271–277. doi:10.1016/j.cognition.2013.11.001

Poliakoff, E., Beach, N., Best, R., Howard, T., & Gowen, E. (2013). Can looking at a hand make your skin crawl? Peering into the uncanny valley for hands. *Perception*, *42*, 998–1000. doi: 10.1068/p7569

Proulx, T., Inzlicht, M., & Harmon-Jones, E. (2012). Understanding all inconsistency compensation as a palliative response to violated expectations. *Trends in Cognitive Sciences*, *16*, 285–291. doi:10.1016/j.tics.2012.04.002

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria. Retrieved from https://www.R-project.org

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (2nd ed.)*. Thousand Oaks, CA: Sage.

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

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience (SCAN), 7*, 413–422. doi:10.1093/scan/nsr025

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

Scott, R. (Director). (1982). *Blade Runner* [Movie]. United States: Warner Bros.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling (2nd ed.).* London: Sage.

Steckenfinger, S. A., & Ghazanfar, A. A. (2009). Monkey visual behavior falls into the uncanny valley. *Proceedings of the National Academy of Sciences, 106*, 18362–18366. doi:10.1073/pnas.0910063106

Tinwell, A., Grimshaw, M., & Williams, A. (2010). Uncanny behaviour in survival horror games. *Journal of Gaming & Virtual Worlds*, *2*, 3–25. doi:10.1386/jgvw.2.1.3_1

Thompson, M. M., Naccarato, M. E., Parker, K. C., & Moskowitz, G. (2001). The personal need for structure and personal fear of invalidity measures: Historical perspectives, current applications, and future directions. In G. Moskowitz (Ed.), *Cognitive social psychology:*

The Princeton symposium on the legacy and future of social cognition (pp. 19–39). Mahwah, NJ: Erlbaum.

Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology*, *19*, 393–407. doi:10.1037/gpr0000056

West, B.T., Welch, K. B., & Gałecki, A. T. (2015). *Linear mixed models: A practical guide using statistical software (2nd ed.).* Boca Raton, FL: CRC Press.

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*, 2020–2045. doi: 10.1037/xge0000014

Wingfield, N. (2015, March 19). Amazon wins approval to test delivery drones outdoors. *The New York Times.* Retrieved from http://www.nytimes.com/

Yamada, Y., Kawabe, T., & Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *Japanese Psychological Research, 55*, 20–32. doi:10.1111/j.1468-5884.2012.00538.x

Footnotes

[1] For nested (noncrossed) two-level data, Geldhof et al. (2014) suggested calculating between-persons alpha across all repeated measures, using a two-level multilevel model (i.e., judgments nested in persons). However, on the between-persons level, this method was not applicable to the data from Experiment 1 because each person saw a different, fixed subset that included three out of nine stimuli, and the stimuli differed systematically in eeriness and human likeness. Therefore, we calculated (between-persons) Cronbach's alphas separately for each of the nine stimuli.

[2] To account for systematic variation between stimulus sets in Experiment 2, we included a random effect for both persons and stimulus sets in this model. That is, instead of a two-level model for nested data as in Geldhof et al.'s (2014) examples, a cross-classified model (with participants and stimulus sets as crossed random effects) was specified to estimate the variance components and calculate alpha on the judgments level and on the between-persons level.

[3] As a measure of effect size, we calculated the proportional reduction in the Level-2 variances of random effects across participants (i.e., the proportional reduction in the variances of u_{0i}, u_{1i}, and u_{2i}) when the PNS subscales were added to the model (quasi-R^2 for Level 2; see e.g., Snijders & Bosker, 2012). The proportional reduction in variance of the intercept was .04, the proportional reduction in variance of the linear slope term for human likeness was .001, and the proportional reduction in variance of the quadratic slope term for human likeness was .06.

Table 1

Descriptive Statistics for the Study Variables in Experiment 1 and Experiment 2

Variable	Experiment 1		Experiment 2	
	M	*SD*	*M*	*SD*
Eeriness	−4.11	1.74	−3.50	2.00
Subjective HL	3.46	2.09	3.75	2.23
PNS total scale	3.70	0.63	3.78	0.73
PNS-DFS	3.99	0.77	4.03	0.93
PNS-RLS	3.53	0.71	3.63	0.77
Extraversion	3.63	0.64	—	—
Neuroticism	2.64	0.62	—	—

Note. HL = Human likeness. PNS = Personal Need for Structure. PNS-DFS = PNS subscale Desire for Structure. PNS-RLS = PNS subscale Response to Lack of Structure. Eeriness is the transformed eeriness variable (possible range from −7 to −1, with lower [i.e., larger negative] scores representing greater eeriness).

Table 2

Results of Multivariate Multilevel Models Predicting Eeriness in Experiment 1

	Fixed				Random (participants)	
	Coef.	Est.	*SE*	*z*	Coef.	*SD*
Model 0						
Dum.Robot	γ_{10}	−4.00	0.10		u_{1i}	1.51
Dum.Android	γ_{20}	−5.51	0.08		u_{2i}	1.15
Dum.Human	γ_{30}	−2.83	0.09		u_{3i}	1.38
Model 1						
Intercept	γ_{00}	−5.51	0.08		u_{0i}	1.15
Dum.Robot	γ_{10}	1.51	0.13	11.99***	u_{1i}	1.89
Dum.Human	γ_{20}	2.68	0.11	23.61***	u_{2i}	1.71
Model 2						
Intercept	γ_{00}	−5.50	0.08		u_{0i}	1.13
Dum.Robot	γ_{10}	1.46	0.11	13.15***	u_{1i}	1.67
Dum.Human	γ_{20}	2.65	0.10	27.32***	u_{2i}	1.46
Eff.Set2	γ_{30}	−0.29	0.10	−2.80**		
Eff.Set3	γ_{40}	0.30	0.11	2.84**		
Dum.Robot ×Eff.Set2	γ_{50}	0.53	0.17	3.10**		
Dum.Robot ×Eff.Set3	γ_{60}	0.40	0.17	2.37*		
Dum.Human × Eff.Set2	γ_{70}	−0.16	0.16	−0.98		
Dum.Human × Eff.Set3	γ_{80}	−0.86	0.16	−5.32***		
Model 3						
Intercept	γ_{00}	−5.50	0.08		u_{0i}	1.13
Dum.Robot	γ_{10}	1.46	0.11	13.15***	u_{1i}	1.67
Dum.Human	γ_{20}	2.65	0.10	27.36***	u_{2i}	1.45
Eff.Set2	γ_{30}	−0.29	0.11	−2.76**		
Eff.Set3	γ_{40}	0.30	0.11	2.82**		
Dum.Robot ×Eff.Set2	γ_{50}	0.52	0.17	3.06**		
Dum.Robot ×Eff.Set3	γ_{60}	0.40	0.17	2.37*		
Dum.Human × Eff.Set2	γ_{70}	−0.17	0.16	−1.05		
Dum.Human × Eff.Set3	γ_{80}	−0.85	0.16	−5.25***		
RLS	γ_{01}	−0.11	0.12	−0.94		
DFS	γ_{02}	0.06	0.11	0.57		
Dum.Robot × RLS	γ_{11}	0.02	0.18	0.12		
Dum.Robot × DFS	γ_{12}	−0.06	0.17	−0.34		
Dum.Human × RLS	γ_{21}	−0.04	0.16	−0.28		
Dum.Human × DFS	γ_{22}	−0.08	0.14	−0.56		

Note. Analyses were conducted in Mplus (which reports z values as the test statistic). For Model 0 (i.e., the baseline model), we do not report z statistics because they would be meaningless (test of the null hypothesis that mean eeriness of a stimulus is zero in the population). In Models 1 to

3, the reference category was the android stimulus. Coef. = Coefficient from model equations in the text. Dum.Robot = Dummy variable for robot stimulus. Dum.Android = Dummy variable for android stimulus. Dum.Human = Dummy variable for human stimulus. Eff.Set2 = Effect-coded indicator variable for stimulus set 2. Eff.Set3 = Effect-coded indicator variable for stimulus set 3. RLS = Response to Lack of Structure. DFS = Desire for Structure.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 3

Correlations between Personality Variables in Experiments 1 and 2

	1	2	3	4	5
1. PNS total scale		.82***	.92***	—	—
2. PNS-DFS	.78***		.52***	—	—
3. PNS-RLS	.92***	.48***		—	—
4. Extraversion	−.30***	−.07	−.38***		—
5. Neuroticism	.37***	.13	.43***	−.29***	

Note. PNS = Personal Need for Structure. PNS-DFS = PNS subscale Desire for Structure. PNS-RLS = PNS subscale Response to Lack of Structure. Correlations between variables in Experiment 1 are depicted below the diagonal, and correlations between variables in Experiment 2 are depicted above the diagonal. *N* persons (Experiment 1) = 226. *N* persons (Experiment 2) = 316.

*** $p < .001$.

Table 4

Results of Mixed Model Analyses Predicting Eeriness by Manipulated Human Likeness in Experiment 2

	Fixed					Random			
						Participants		Stimulus sets	
	Coef.	Est.	*SE*	*t*	*df*	Coef.	*SD*	Coef.	*SD*
Model 1									
Intercept	γ_{00}	−4.34				u_{0i}	1.07	v_{0j}	0.24
Manip. HL	γ_{10}	0.46	0.07	6.15***	8	u_{1i}	0.24	v_{1j}	0.20
Manip. HL^2	γ_{20}	0.21	0.02	8.55***	7	u_{2i}	0.08	v_{2j}	0.07
Manip. HL^3	γ_{30}	−0.02	0.00	−8.99***	7,878				
Model 2									
Intercept	γ_{00}	−4.34				u_{0i}	1.05	v_{0j}	0.24
Manip. HL	γ_{10}	0.46	0.07	6.15***	8	u_{1i}	0.24	v_{1j}	0.20
Manip. HL^2	γ_{20}	0.21	0.02	8.56***	7	u_{2i}	0.08	v_{2j}	0.07
Manip. HL^3	γ_{30}	−0.02	0.00	−9.00***	7,876				
RLS	γ_{01}	−0.32	0.10	−3.33***	313				
DFS	γ_{02}	0.21	0.08	2.75**	313				
Manip. HL× RLS	γ_{11}	0.08	0.04	2.21*	1,600				
Manip. HL^2 × RLS	γ_{21}	0.03	0.01	2.87**	313				
Manip. HL^3 × RLS	γ_{31}	−0.00	0.00	−1.60	7,878				
Manip. HL × DFS	γ_{12}	−0.01	0.03	−0.37	1,598				
Manip. HL^2 × DFS	γ_{22}	−0.02	0.01	−3.14**	313				
Manip. HL^3 × DFS	γ_{32}	−0.00	0.00	−0.36	7,875				
Model 3									
Intercept	γ_{00}	−4.34				u_{0i}	1.06	v_{0j}	0.24
Manip. HL	γ_{10}	0.46	0.07	6.15***	8	u_{1i}	0.24	v_{1j}	0.20
Manip. HL^2	γ_{20}	0.21	0.02	8.55***	7	u_{2i}	0.08	v_{2j}	0.07
Manip. HL^3	γ_{30}	−0.02	0.00	−9.00***	7,877				
RLS	γ_{01}	−0.18	0.08	−2.19*	314				
Manip. HL× RLS	γ_{11}	0.07	0.03	2.36*	1,605				
Manip. HL^2 × RLS	γ_{21}	0.01	0.01	1.42	314				
Manip. HL^3 × RLS	γ_{31}	−0.01	0.00	−2.10*	7,879				
Model 4									
Intercept	γ_{00}	−4.34				u_{0i}	1.06	v_{0j}	0.24
Manip. HL	γ_{10}	0.46	0.07	6.15***	8	u_{1i}	0.24	v_{1j}	0.20
Manip. HL^2	γ_{20}	0.21	0.02	8.55***	7	u_{2i}	0.08	v_{2j}	0.07
Manip. HL^3	γ_{30}	−0.02	0.00	−9.00***	7,877				
DFS	γ_{01}	0.08	0.07	1.17	314				
Manip. HL× DFS	γ_{11}	0.02	0.03	0.91	1,597				
Manip. HL^2 × DFS	γ_{21}	−0.01	0.01	−1.90$^+$	314				
Manip. HL^3 × DFS	γ_{31}	−0.00	0.00	−1.41	7,877				

Note. Mixed model analyses were computed in R (using the packages lme4 and lmerTest). Coef. = Coefficient from model equations in the text. Manip. HL = Manipulated human likeness (steps of the morphing continuum). RLS = Response to Lack of Structure. DFS = Desire for Structure. $^{+}$ $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table 5

Results of Mixed Model Analyses Predicting Eeriness by Subjective Human Likeness in Experiment 2

	Fixed					Random			
						Participants		Stimulus sets	
	Coef.	Est.	*SE*	*t*	*df*	Coef.	*SD*	Coef.	*SD*
Model 1									
Intercept	γ_{00}	−4.27				u_{0i}	1.06	v_{0j}	0.33
Subj. HL	γ_{10}	0.39	0.06	6.46***	10	u_{1i}	0.28	v_{1j}	0.16
Subj. HL^2	γ_{20}	0.17	0.01	12.75***	12	u_{2i}	0.10	v_{2j}	0.03
Subj. HL^3	γ_{30}	−0.02	0.00	−5.96***	3,102				
Model 2									
Intercept	γ_{00}	−4.27				u_{0i}	1.05	v_{0j}	0.33
Subj. HL	γ_{10}	0.40	0.06	6.48***	10	u_{1i}	0.28	v_{1j}	0.16
Subj. HL^2	γ_{20}	0.17	0.01	12.83***	11	u_{2i}	0.10	v_{2j}	0.03
Subj. HL^3	γ_{30}	−0.02	0.00	−6.10***	3,117				
RLS	γ_{01}	−0.29	0.10	−3.02**	300				
DFS	γ_{02}	0.24	0.08	2.97**	302				
Subj. HL × RLS	γ_{11}	0.07	0.04	1.83^{+}	1,183				
Subj. HL^2 × RLS	γ_{21}	0.02	0.01	1.79^{+}	342				
Subj. HL^3 × RLS	γ_{31}	−0.01	0.00	−1.71^{+}	3,466				
Subj. HL × DFS	γ_{12}	−0.02	0.03	−0.69	1,172				
Subj. HL^2 × DFS	γ_{22}	−0.03	0.01	−3.00**	341				
Subj. HL^3 × DFS	γ_{32}	0.01	0.00	1.79^{+}	3,283				
Model 3									
Intercept	γ_{00}	−4.27				u_{0i}	1.06	v_{0j}	0.33
Subj. HL	γ_{10}	0.40	0.06	6.47***	10	u_{1i}	0.28	v_{1j}	0.16
Subj. HL^2	γ_{20}	0.17	0.01	12.75***	12	u_{2i}	0.10	v_{2j}	0.03
Subj. HL^3	γ_{30}	−0.02	0.00	−5.98***	3,079				
RLS	γ_{01}	−0.14	0.08	−1.73^{+}	304				
Subj. HL× RLS	γ_{11}	0.06	0.03	1.72^{+}	1,208				
Subj. HL^2 × RLS	γ_{21}	0.00	0.01	0.28	344				
Subj. HL^3 × RLS	γ_{31}	−0.00	0.00	−0.92	3,447				
Model 4									
Intercept	γ_{00}	−4.27				u_{0i}	1.06	v_{0j}	0.33
Subj. HL	γ_{10}	0.40	0.06	6.48***	10	u_{1i}	0.28	v_{1j}	0.16
Subj. HL^2	γ_{20}	0.17	0.01	12.82***	12	u_{2i}	0.10	v_{2j}	0.03
Subj. HL^3	γ_{30}	−0.02	0.00	−6.07***	3,100				
DFS	γ_{01}	0.11	0.07	1.63	307				
Subj. HL× DFS	γ_{11}	0.01	0.03	0.30	1,199				
Subj. HL^2 × DFS	γ_{21}	−0.02	0.01	−2.44*	340				
Subj. HL^3 × DFS	γ_{31}	0.00	0.00	1.04	3,255				

Note. Mixed model analyses were computed in R (using the packages lme4 and lmerTest). Coef. = Coefficient from model equations in the text. Subj. HL = Subjective human likeness ratings. RLS = Response to Lack of Structure. DFS = Desire for Structure.

$^{+}$ $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

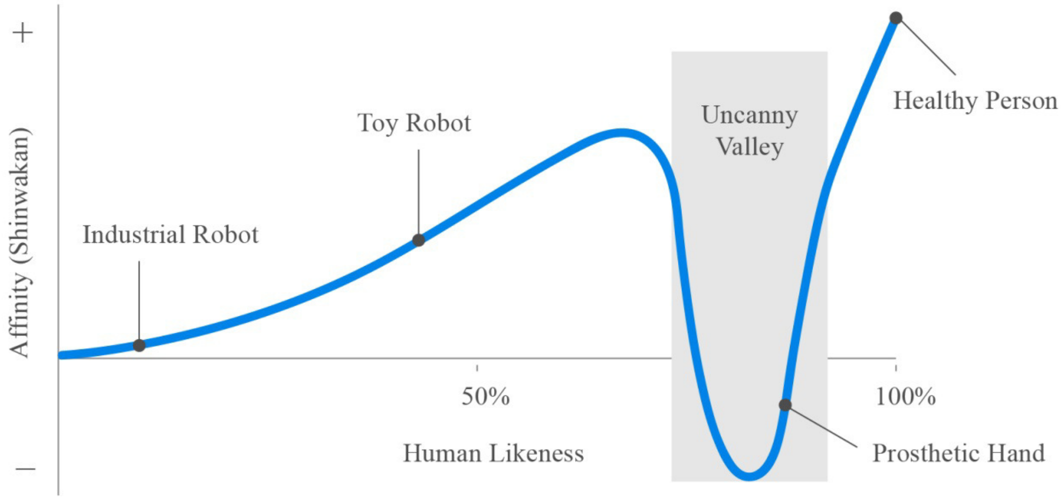

Figure 1. The nonlinear relationship between the human likeness of stimuli and users' affinity. The uncanny valley proper is emphasized with a grey color. Adapted from MacDorman (2005).

	Human likeness		
Stimulus Set	Robot	Android	Human
1			
2	Photo of Robot Bandit		
3			

Figure 2. Stimuli used in Experiment 1 (three degrees of human likeness for three stimulus sets). The robot images (from top to bottom row) show the robots called Aila (DFKI Robotics Innovation Center), Bandit (USC Interaction Lab and BlueSky Robotics), and Roboy (AI Lab, University of Zurich), taken from the IEEE robot collection (Institute of Electrical and Electronic Engineers, 2012). The photo of Bandit is not depicted because we were not able to obtain the permission to reprint the photo. The photo can be found at http://robotsapp.spectrum.ieee.org. The female human image depicts the actress Keira Knightley and was retrieved via a Google Image search. The male human images were taken from the Karolinska Directed Emotional Faces database (Lundqvist et al., 1998; image IDs AM10NES, BM09NES).

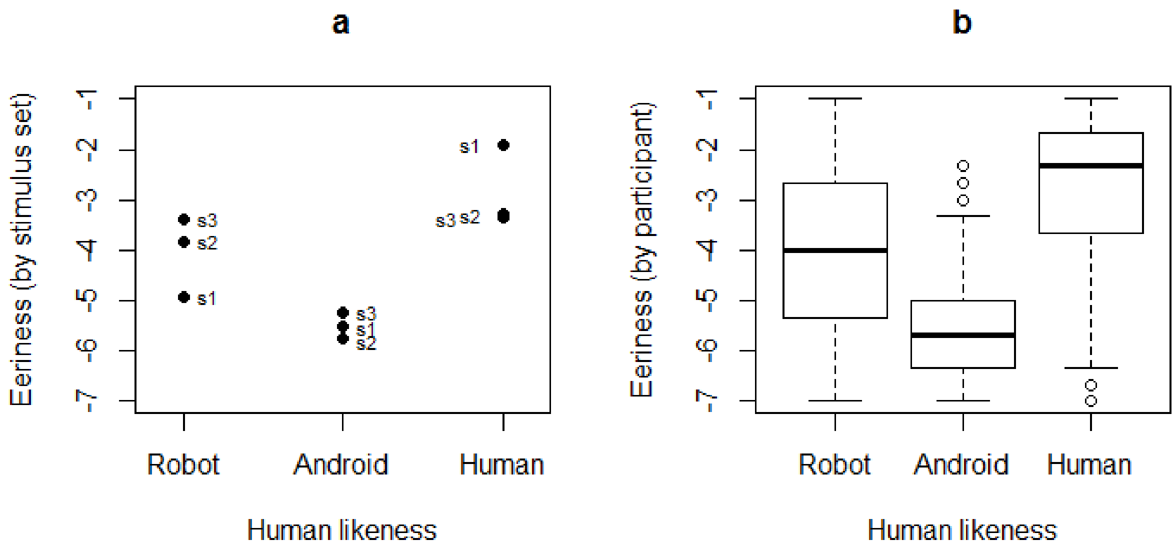

Figure 3. Eeriness as a function of manipulated human likeness in Experiment 1: a) By-stimulus set analyses (mean eeriness ratings for each of the three stimulus sets averaged across participants, with stimulus set number placed next to each mean); b) By-participant analysis (eeriness ratings of each participant).

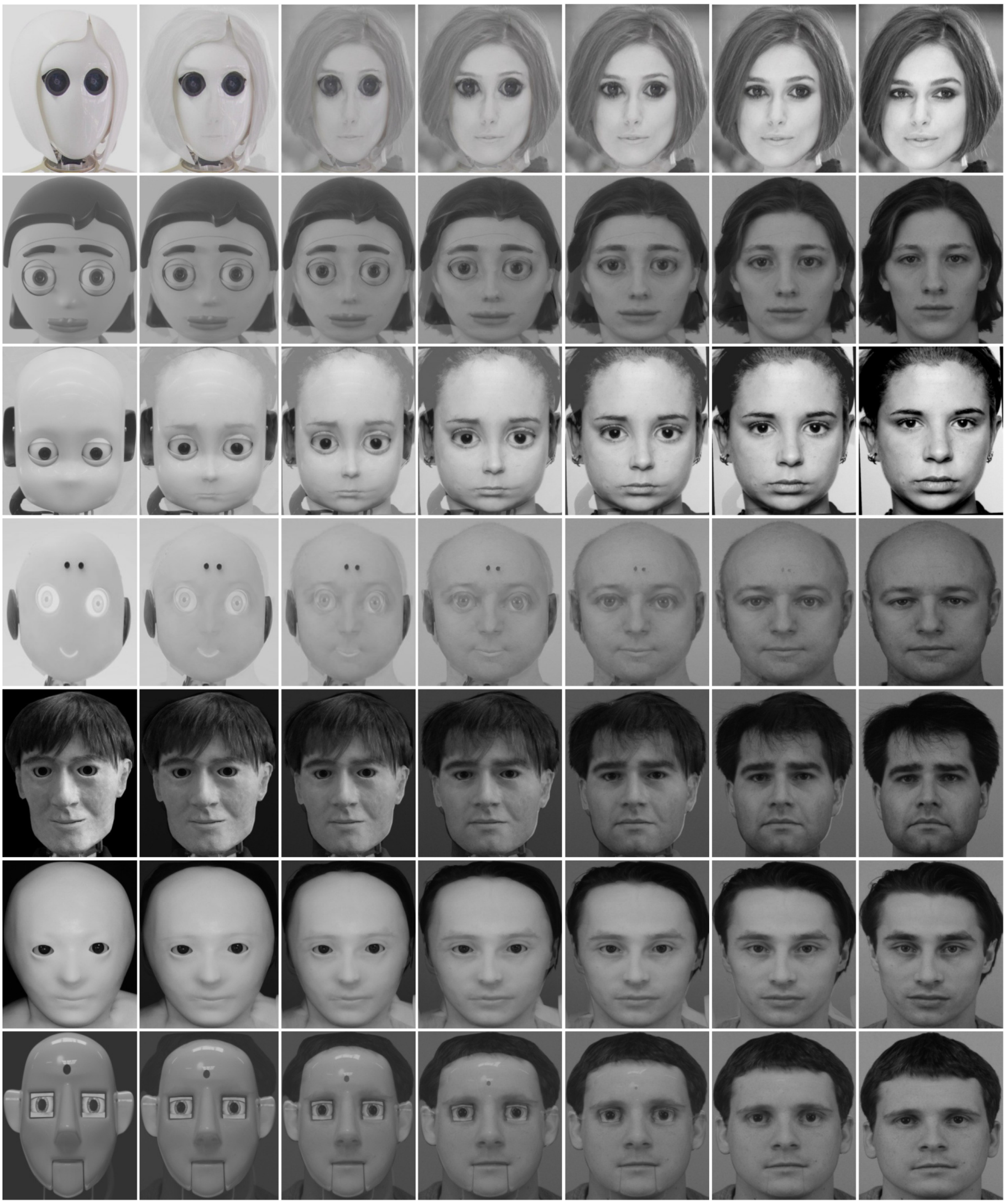

Figure 4. Stimuli used in Experiment 2. Rows represent the eight stimulus sets, columns represent the seven steps of the morphing continuum from robot (left) to human (right). The robot images (left column, from top to bottom row) show the robots Aila (DFKI Robotics Innovation Center), Flobi (University of Bielefeld, Germany), iCub (RoboCub Consortium, Bar Ilan

University, Israel), Roboy (AI Lab, University of Zurich, Switzerland), Roman (Robotics Research Lab, University of Kaiserslautern, Germany), Telenoid™ (Ishiguro Lab at Osaka University, Japan), and RoboThespain (Engineered Arts Ltd, United Kingdom). Telenoid™ has been developed by Osaka University and Hiroshi Ishiguro Laboratories, Advanced Telecommunications Research Institute International (ATR). The male human images were selected from the Karolinska Directed Emotional Faces database (Lundqvist et al, 1998; image IDs AM10NES, AM18NES, BM09NES, AM30NES, BM25NES, AM31NES), and the two female images were extracted from a Google-Image search. The eighth stimulus set is not depicted. It consisted of an image of the robot Bandit (USC Interaction Lab and BlueSky Robotics), a male human image from the KDEF (ID AM10NES) and five morphs. The reason for not depicting this stimulus set is that we were not able to obtain the permission to reprint the photo of robot Bandit. It can be found at http://robotsapp.spectrum.ieee.org.

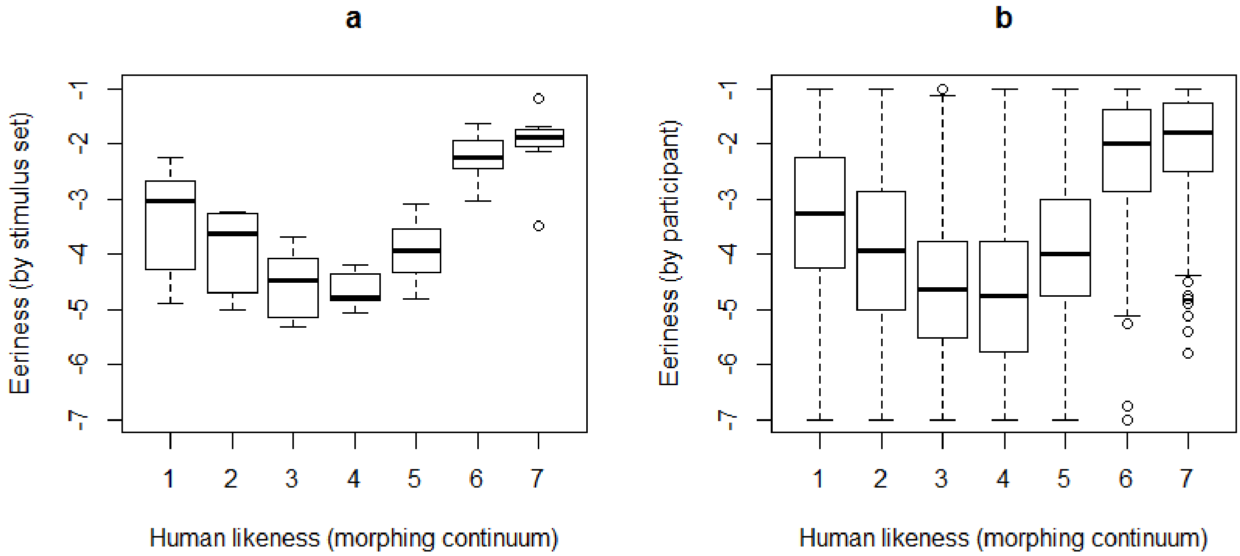

Figure 5. Eeriness as a function of manipulated human likeness (steps of the morphing continuum) in Experiment 2: a) By-stimulus set analyses (eeriness ratings for each stimulus set averaged across participants); b) By-participant analysis (eeriness ratings of each participant averaged across stimuli).

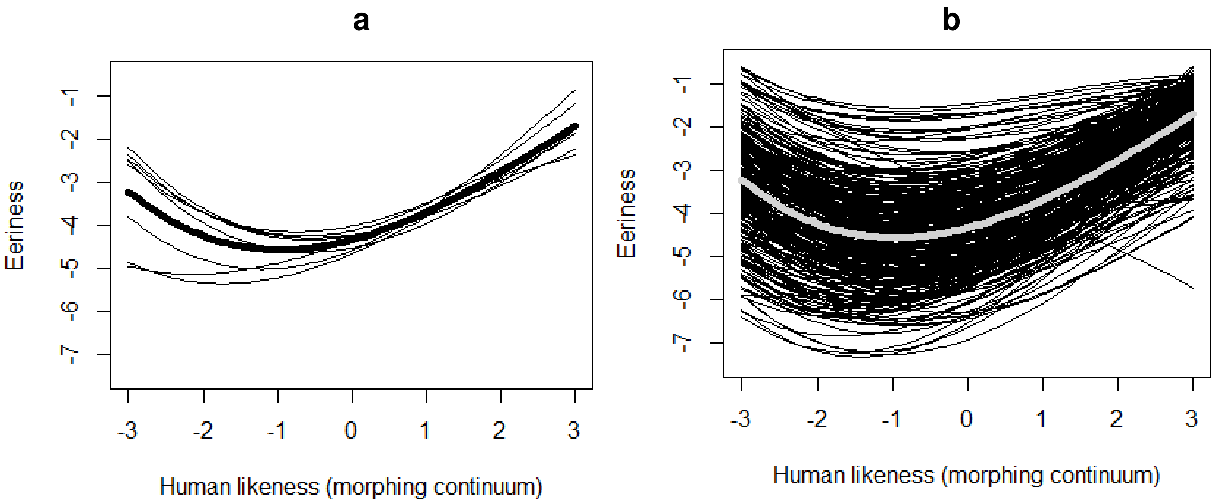

Figure 6. Curvilinear relationship between manipulated human likeness (centered at the midpoint of the continuum) and eeriness as estimated from the mixed model in Experiment 2: a) Stimulus-set-specific curves (depicting random effects across stimulus sets) and the average curve (across participants and stimulus sets) in bold; b) participant-specific curves (depicting random effects across participants) and the average curve (across participants and stimulus sets) in bold grey.

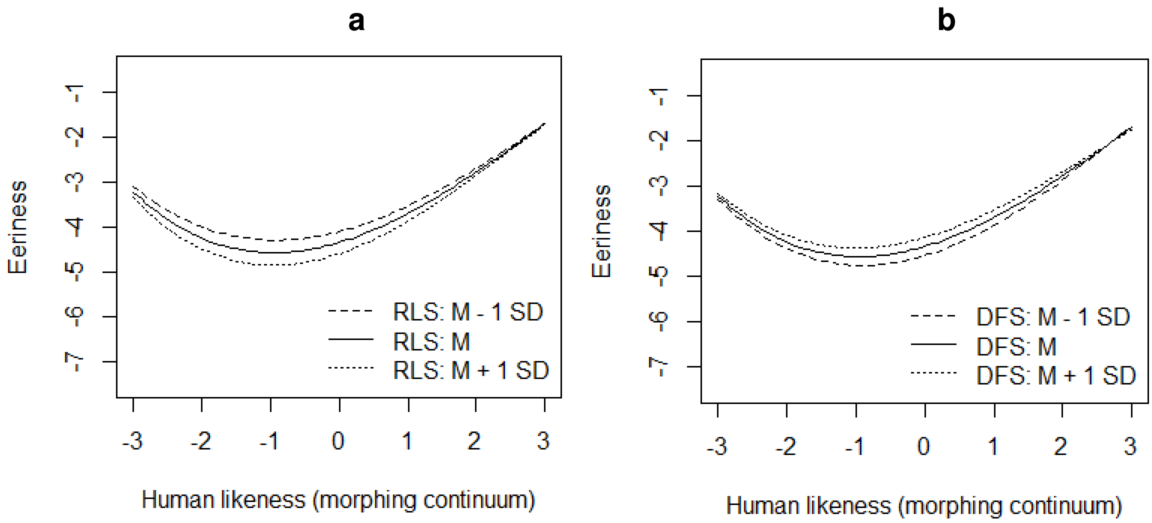

Figure 7. PNS subscales as moderators of the relationship between manipulated human likeness and eeriness as estimated from the mixed model (Model 2 in Table 4) in Experiment 2: a) Moderator effect of PNS subscale Response to Lack of Structure (RLS); b) Moderator effect of PNS subscale Desire for Structure (DFS).