Review Article

Human-Like Robots and the Uncanny Valley

A Meta-Analysis of User Responses Based on the Godspeed Scales

Martina Mara[1], Markus Appel[2], and Timo Gnambs[3]

[1]LIT Robopsychology Lab, Johannes Kepler University Linz, Austria
[2]Psychology of Communication and New Media, University of Würzburg, Germany
[3]Leibniz Institute for Educational Trajectories (LIfBi), University of Bamberg, Germany

Abstract: In the field of human-robot interaction, the well-known uncanny valley hypothesis proposes a curvilinear relationship between a robot's degree of human likeness and the observers' responses to the robot. While low to medium human likeness should be associated with increased positive responses, a shift to negative responses is expected for highly anthropomorphic robots. As empirical findings on the uncanny valley hypothesis are inconclusive, we conducted a random-effects meta-analysis of 49 studies (total N = 3,556) that reported 131 evaluations of robots based on the Godspeed scales for anthropomorphism (i.e., human likeness) and likeability. Our results confirm more positive responses for more human-like robots at low to medium anthropomorphism, with moving robots rated as more human-like but not necessarily more likable than static ones. However, because highly anthropomorphic robots were sparsely utilized in previous studies, no conclusions regarding proposed adverse effects at higher levels of human likeness can be made at this stage.

Keywords: uncanny valley, humanoid robot, anthropomorphism, likeability, meta-analysis

https://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000486 - Thursday, March 28, 2024 3:48:59 PM - IP Address:44.204.94.166

When people think of robots, they usually have an image of a human-like machine in their minds: an apparatus with arms, legs, and a head, covered in metal or possibly silicone skin (see Cave et al., 2020; Mara et al., 2020). Even though such robots hardly, if at all, exist in our everyday lives, media reports about engineering advancements and science fiction stories about the – sometimes more, sometimes less peaceful – relationship between humans and their robotic counterparts have long made us wonder what it would be like if humanoid machines were really among us. Given the diffuse mental pictures many people have about robots, representative survey data show that many people are skeptical regarding their use in everyday life (e.g., Gnambs, 2019; Gnambs & Appel, 2019). One of the most popular conceptual frameworks to speculate about human responses to human-like robots is the uncanny valley hypothesis (Mori, 1970). Its central proposition is that increasing anthropomorphism (i.e., human likeness) in artificial characters does not necessarily go hand in hand with increasing likeability but will result in negative responses when the degree of human resemblance is very high, yet not perfect. Over the past decade, the number of empirical investigations of human-robot relationships and determinants of robot acceptance has steadily increased, many of which have dealt with potentially aversive reactions to human-like machines. However, due to inconsistent empirical evidence, the existence of the uncanny valley effect and the conditions under which it is more or less pronounced are a matter of debate (see Kätsyri et al., 2015; Wang et al., 2015; Zhang et al., 2020). Given the great popularity of the uncanny valley hypothesis, it is surprising that its basic propositions still lack systematic empirical corroboration. We address this gap by conducting the first meta-analytic test of the curvilinear relationship between the human likeness and the likeability of robots as proposed by Mori (1970).

Human-Like Robots

From mythological figures such as the Golem to modern-day science fiction, stories about artificial replications of the human species were told throughout history. Starting in the 18th century, there have also been attempts to physically create human-like machines. Around the first industrial revolution, watchmakers and mechanical engineers constructed life-sized automatons in the shape of adult humans that appeared as if they could write, draw, or play chess (see Voskuhl, 2013). When the term "robot" was first ever used in the context of the 1920 theater play "Rossum's Universal Robots" (Čapek, 1920/2001), it was also

© 2022 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY-NC-ND 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000486 - Thursday, March 28, 2024 3:48:59 PM - IP Address:44.204.94.166

human-like automata that were shown on stage. Today, the imitation of the human body and mind constitutes an objective that is being pursued in subdisciplines of robotics and artificial intelligence. While the number of functional human-like robots is still quite small to date, some robotics labs specialize in developing human-like autonomous machines that can serve entertainment purposes (Johnson et al., 2016), answer questions to customers (Pandey & Gelin, 2018), facilitate telepresence (Ogawa et al., 2011), assist in healthcare (Yoshikawa et al., 2011), act as sex toys (Döring et al., 2020), or are used for research into human behavior and bodily functions (Hoffmann & Pfeifer, 2018). Depending on how easily they can be distinguished from real people, human-like robots are typically referred to either as *humanoids* or *androids*. Humanoid robots are easily recognized as robots by their overall mechanical look, even though they usually possess a head, torso, arms, and sometimes legs. In contrast, android robots are intended to mimic human appearance as realistically as possible, emphasized for example, by silicone skin, clothing, wigs, or highly realistic details such as eyelashes (see Ishiguro, 2016).

The Uncanny Valley Hypothesis

Many years before robotics could even draw near the development of real android robots, Japanese roboticist Masahiro Mori introduced the hypothetical model of the uncanny valley (Mori, 1970). Initially intended more as a philosophical contribution than a blueprint for empirical research, after many years of little attention, the uncanny valley turned into a much-discussed and much-studied concept in the past two decades. The popular uncanny valley graph (Figure 1), which was originally based only on Mori's personal experience and conjecture, proposes a nonlinear relationship between the human likeness of an artificial figure, for example, of a robot, and the valence it elicits in observers. Mori suggested that within a spectrum of a generally low to medium degree of visual anthropomorphism, increasing levels of human likeness are associated with increasing acceptance and likeability. Observers should therefore sympathize more strongly with a slightly humanoid robot than, for example, with a swivel-arm robot from the industry. However, after a first positive peak of the curve along the human likeness continuum, this effect should reverse as soon as a rather high level of nearly realistic human likeness is obtained. At this point, acceptance is expected to drop, and the android should evoke a negative and irritating feeling of uncanniness (eeriness, creepiness). As an inherent property of animated entities, motion is moreover assumed to moderate the uncanny valley effect, with moving robots eliciting more pronounced reactions than static objects (or static pictures of moving objects). Therefore, a moving, highly human-like android robot should be perceived as less likable than the corresponding still artifact. Ultimately, on the right side of the uncanny valley, the likeability curve is expected to go up again when a robot's design is so perfectly realistic that it becomes indistinguishable from a real person. At the upper end of the human likeness continuum, at which the real human constitutes the endpoint, the valence of associated affect and cognition should then reach a second positive peak (Mori, 1970; Mori et al., 2012).

Different perceptual, cognitive, or evolutionary explanations have been proposed to underly the uncanny valley phenomenon, including assumptions related to categorical uncertainty, difficulties in the configural processing of human-like artifacts, threat avoidance, or the role of android robots as salient reminders of human mortality (see Diel & MacDorman, 2021; Wang et al., 2015, for an overview of suggested mechanisms).

Research on the Uncanny Valley

Compared to other scientific fields, research on the uncanny valley is characterized by a great diversity of involved disciplines, ranging from robotics, computer science, and virtual reality to animation, design, philosophy, communication science, and psychology. It, therefore, comes as no great surprise that the available studies exhibit considerable methodological heterogeneity. While, for example, a number of researchers investigated the uncanny valley by presenting study participants with physical humanoid or android robots (e.g., Bartneck, Kanda, et al., 2009; Mara & Appel, 2015) or with media representations of actually existent robots (e.g., Kim et al., 2020), other scholars focused on computer-generated stimuli such as virtual faces and avatars (e.g., Kätsyri et al., 2019; Stein & Ohler, 2017) or self-created image morphs (e.g., Lischetzke et al., 2017). Independent of the visual appearance of robots, a more recent branch of uncanny valley research also deals with aversive reactions to purely behavioral human likeness, partly relying on textual descriptions of robots as stimuli (e.g., Appel et al., 2020). Different approaches also prevail in the operationalization of central variables and associated measurements. Single-item self-reports appear to be a common means in research on user responses to human-like robots. Regarding validated multi-item scales for investigations of the uncanny valley, it is, in particular, the Godspeed questionnaire by Bartneck, Kulić, and colleagues (2009) that can be regarded as a dominant instrument for the assessment of robot anthropomorphism (representing the x-axis in Figure 1) and robot likeability (representing the y-axis in Figure 1) (see Weiss & Bartneck, 2015). Another multi-item measure, the uncanny valley indices by Ho and MacDorman (2010, 2017), has been utilized in few studies.

© 2022 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY-NC-ND 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0)

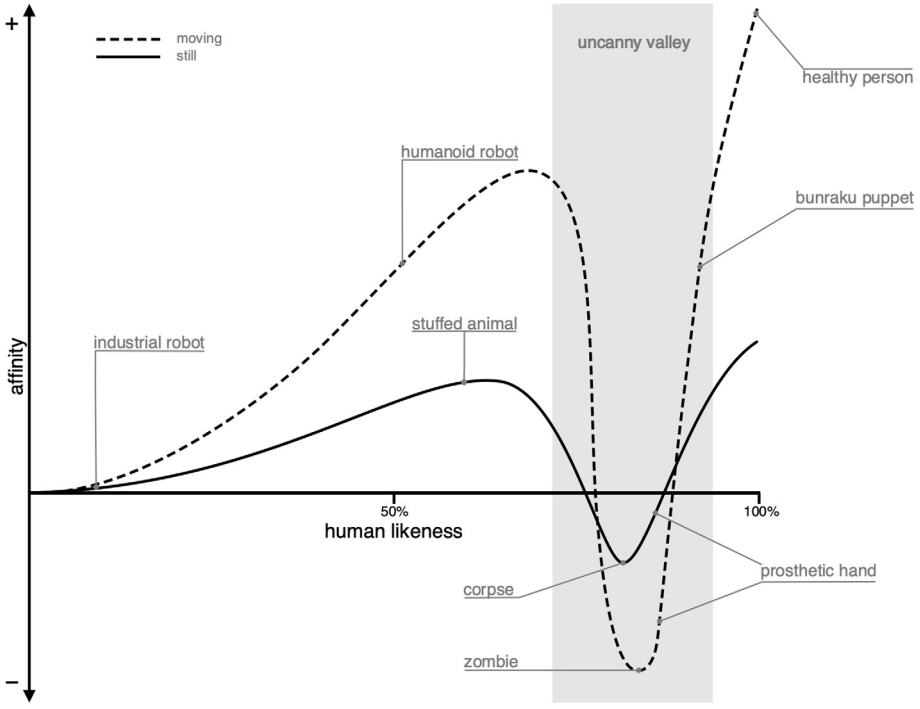


Figure 1. Uncanny Valley Hypothesis (after Mori, 1970).

Empirical support for the idea of the uncanny valley itself has been inconsistent. While results from some studies provide evidence for Mori's propositions (e.g., Mathur & Reichling, 2016) or found partial support (e.g., Bartneck et al., 2007), others failed to reveal a drop in acceptance for highly anthropomorphic machines (e.g., Bartneck, Kanda, et al., 2009) or even revealed an additional uncanny valley along the human likeness continuum (Kim et al., 2020). A literature review (Kätsyri et al., 2015) concluded that a bulk of studies supported a linear increase in affinity for more human-like robots, while evidence for nonlinear uncanny valley effects was scarce. Similarly, the assumption that robot motion should result in stronger uncanny valley effects (see Figure 1) was rarely corroborated (Piwek et al., 2014; Thompson et al., 2011). So far, a quantitative summary of uncanny valley effects is sorely missing.

The Present Study

One factor that contributes to the heterogeneity of study results on the uncanny valley might be the use of unstandardized measurements of the core constructs that exhibit unknown reliability and validity (see Wang et al., 2015). Therefore, the present meta-analysis focuses on the multi-item Godspeed questionnaire (Bartneck, Kulić, et al., 2009) that constitutes a widely used instrument for the assessment of both anthropomorphism and likeability in human-robot interaction research. It can be used to map values on both the *x*-axis and the *y*-axis of the uncanny valley graph. In the interest of ecological validity, we furthermore decided to only include studies in which participants were presented with actual robotic systems or media representations of such. To examine the central propositions of the uncanny valley effect as suggested by Mori (1970) in Figure 1, we hypothesized that (a), overall, with increasing human likeness attributed to a robot, it will be rated more positively (i.e., higher likeability).[1] Moreover, (b) the association between human likeness and likeability should show a nonlinear relationship, leading to (c) an inverted U-shaped function and thus a sharp decline of likeability ratings for highly but not perfectly anthropomorphic robots. Furthermore, (d) a second turning point at the end of the inverted U-shape at the bottom of the valley was expected to lead to more positive ratings for the most human-like robotic agents that are (nearly) indistinguishable from humans. Finally, we assumed (e) robot motion to have a moderating role because Mori (1970) speculated that motion, as an inherent property of animated objects, should amplify the uncanny valley effect.

[1] Nonlinear prediction models such as the Uncanny Valley hypothesis might exhibit an average linear trend, which is then specified in detail by nonlinear associations between the focal variables.

https://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000486 - Thursday, March 28, 2024 3:48:59 PM - IP Address:44.204.94.166

© 2022 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY-NC-ND 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000486 - Thursday, March 28, 2024 3:48:59 PM - IP Address:44.204.94.166

Method

Literature Search and Study Selection

In January 2021, we performed a literature search for studies in which at least one robot was evaluated with the help of the Godspeed questionnaire by identifying articles in Google Scholar, citing Bartneck, Kulić, and colleagues (2009). Initial search results provided 1,330 potentially relevant publications. After screening the titles, abstracts, and method sections of these articles, 95 records were subjected to detailed evaluations. To be included in the meta-analysis, a study had to meet the following criteria. First, it had to have administered the anthropomorphism and likeability scales of the Godspeed questionnaire without substantial changes to the item content. However, we considered short forms of the scale if they included at least two items, and we allowed for deviations in the number of response options (from the original 5-point ratings). Second, the respondents interacted with or viewed a real robot, a close reproduction of a real robot, or viewed a photograph or video of a robot. Virtual agents, avatars, morphed images, fictional representations (e.g., drawings, caricatures), or mere verbal descriptions of robots were not considered. No restrictions were applied on the size or the form of the robot to cover technical systems with a broad range of human likeness. Third, the study must have reported means, standard deviations, and sample sizes for both scales or provided information to derive these statistics (e.g., plots). Fourth, the study must have included healthy samples without psychological disorders. Finally, we acknowledged all studies published until December 2020. No restrictions were set on the publication type. After applying these criteria, 49 publications reporting on 93 independent samples were available (see the flow diagram in the Electronic Supplementary Material, ESM 1).

Data Extraction

From each article, we coded the mean, standard deviation, reliability (coefficient alpha), number of administered items, and number of response options for the anthropomorphism and likeability scales. For 19 studies that did not report numeric results, means and standard deviations were approximated from plots (e.g., histograms with standard errors) using the *R* package *metaDigitise* version 1.0.1 (Pick et al., 2019). In case a study reported on multiple robots, we coded each robot separately. In contrast, if different ratings were presented for the total sample and different subgroups (e.g., different experimental conditions), we only coded the results for the total sample (i.e., with the largest sample size). However, if the information was available for different values of the examined moderators (see below), then results for the different subgroups (i.e., whether the robot moved or talked) were coded separately. Additionally, we recorded the name of the evaluated robot, how it was presented (real, photo, video, virtual reality), whether it moved, and whether it communicated (e.g., talked or made sounds). Descriptive information on the sample included the sample size, the mean age of the respondents, the share of females, the country of origin of the participants, and the language of administration. Finally, we noted the publication year and the publication type (journal, proceedings, book chapter, thesis) of each study. All studies were coded by the last author and, independently, by three research assistants. Additionally, the risk of bias for each study was evaluated by two research assistants using eight items of the *Risk of Bias Utilized for Surveys Tool*, a checklist to code quality criteria such as the acceptability of exclusion rates or the sufficiency of sample sizes for primary studies used in meta-analyses (Nudelman & Otto, 2020).

For most coded variables, the interrater reliability (Krippendorff's alpha) indicated good agreement exceeding $\alpha_K \geq .85$ (*Mdn* = .90). However, the codings of the sample sizes ($\alpha_K = .63$) and whether the robot moved ($\alpha_K = .31$) or communicated ($\alpha_K = .66$) were less consistent. The interrater reliability of the risk of bias assessments was good with $\alpha_K = .91$. Discrepancies were solved by the first author. The characteristics of the samples, including the coded statistics, are summarized in ESM 1.

Analysis Plan

Because the uncanny valley hypothesis refers to a nonlinear association between anthropomorphism and likeability, the means of the likeability scale were the focal statistics that were pooled across studies. A random-effects meta-analysis was conducted using the *metafor* software version 2.4-0 (Viechtbauer, 2010) with a restricted maximum likelihood estimator. To account for sampling error, the means were weighted by the inverse of their sampling variances. Because some studies reported more than one evaluation (e.g., obtained for different robots), we estimated a three-level meta-analytic model that acknowledged dependencies between samples using a random-effects structure (see Cheung, 2019; Van den Noortgate et al., 2013). The uncanny valley effect was examined using polynomial meta-regression analyses that predicted likeability ratings from anthropomorphism scores. To model the hypothesized inflection points (see Figure 1) the regression also included higher-order polynomials of the anthropomorphism scores. In sensitivity analyses, we included several additional covariates (e.g., share of female respondents, risk of bias) and repeated the polynomial regression to determine the robustness of the observed effects. Moreover,

© 2022 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY-NC-ND 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000486 - Thursday, March 28, 2024 3:48:59 PM - IP Address:44.204.94.166

we also repeated these analyses, excluding outliers (Viechtbauer & Cheung, 2010) and using robust meta-regression analyses (Hedges et al., 2010) to highlight the generalizability of results against different methodological choices (see Voracek et al., 2019). The homogeneity of the pooled scores was tested using the χ^2-distributed Q-statistic and quantified using I^2 that indicates the percentage of the total variance in observed scores due to random variance. Moderators were evaluated using the χ^2-distributed omnibus test statistic Q_m. The precision of the predicted nonlinear association between anthropomorphism and likeability was determined using a 95% confidence interval. All analyses were conducted in *R* version 4.03 (R Core Team, 2020).

Open Practices

The checklist for the *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (Page et al., 2021) is provided in ESM 1. To foster transparency and reproducibility, we also provide the coding manual, extracted data, computer code, and analysis results at https://osf.io/t9rdk. The meta-analysis was not preregistered.

Results

Description of Meta-Analytic Database

The meta-analytic database included 49 studies that reported on 93 independent samples and included 131 evaluations of robots. Each sample contributed between 1 and 9 (*Mdn* = 1) evaluations of a robot using the Godspeed scales, predominantly in their original form, including five items and 5-point response scales. Both scales exhibited good reliabilities with median coefficient alphas of .86 for anthropomorphism and .89 for likeability. Results of respective reliability generalizations are summarized in ESM 1. Key characteristics of the included samples are also given in Table 1. The sample sizes ranged from 6 to 121 and included a median of 21 respondents. Most samples were from Germany (44%) and the United Kingdom (11%). The median proportion of female participants was 50%. Although the mean age of the samples spanned a broad range from 9 to 68 years, most samples were rather young (*Mdn* = 25 years) and dominated by students or university personnel (79%). Few studies included more diverse groups such as individuals with lower education (Trovato et al., 2015b), children (Meghdari et al., 2018; Shariati et al., 2018), or senior citizens (Rosenthal-von der Pütten et al., 2017). About 55% of studies were published in conference proceedings, while journal articles (33%) were less prevalent. The risk of bias assessments had a median of 3 (on a scale from 0 to 8) and, thus, indicated that many studies exhibited several designs or reporting weaknesses that might have limited the validity of the reported study results to some degree.

Evaluations of Robots

The studied robots came in different forms and sizes, representing a broad range of different models. Most available ratings pertained to the NAO robot by SoftBank Robotics (33%), the iCub robot by the Italian Institute of Technology (8%), and the Pepper robot by SoftBank Robotics (7%). In addition, various custom-built robots were examined, such as the bartender robot JAMES (Foster et al., 2012; Giuliani et al., 2013), the neuro-inspired companion robot NICO (Kerzel et al., 2020), the blessing robot BlessU2 (Löffler et al., 2019), a Sunflower housing robot (Syrdal et al., 2013), or the industrial robot ARMAR-6 (Busch et al., 2019). The distributions of the average anthropomorphism and likeability scores for these robots in Figure 2 highlight two intriguing results. First, the observed anthropomorphism scores ranged between 1.20 and 4.14, and most ratings fell in the lower middle range of possible scores (*Mdn* = 2.61). Thus, human likeness scores in the upper range were scarce. Second, the observed likeability scores ranged between 2.63 and 4.98 (*Mdn* = 3.92). This implies that most robots were rated moderately to very favorably, whereas only a few likeability ratings were in the low range.

However, there were notable differences in these evaluations between different robot models. Therefore, we pooled the anthropomorphism and likeability scores for selected robot models and summarized the meta-analytic estimates in Figure 3. Detailed meta-analytic results, based on calculations in which we used the robot model as a predictor in a meta-regression, are reported in ESM 1. For example, the bartender robot JAMES was rated significantly ($p < .05$) less human-like as compared to the average rating across all robots. In contrast, the iCub robot and Pepper received significantly higher anthropomorphism scores (see Table E2 in ESM 1). A rather similar picture emerged for the pooled likeability ratings. While the bartender robot JAMES was evaluated significantly less likable as compared to the average evaluation, the NAO robot was evaluated significantly more likable. Interestingly, the robot model explained about 20% in anthropomorphism scores, while it only accounted for about 4% in likeability ratings.

Tests of the Uncanny Valley Hypothesis

The association between the two Godspeed scales was examined using meta-regression analyses that predicted the likeability scores from the anthropomorphism ratings.

© 2022 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY-NC-ND 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000486 - Thursday, March 28, 2024 3:48:59 PM - IP Address:44.204.94.166

Table 1. Descriptive statistics for samples included in the meta-analytic database

Variable	*Mdn*/%	Min	Max	Valid	Missing
Sample size	21	6	121	93	0%
Number of evaluations per sample	1	1	9	93	0%
Country of origin				80	14%
Germany	44%				
Italy	5%				
Japan	5%				
The Netherlands	6%				
United Kingdom	11%				
Other	29%				
Publication year	2018	2011	2020	93	0%
Percentage females	50	0	81	89	4%
Mean age	25	9	68	80	14%
Sample type				65	30%
Students/university personnel	79%				
General public	9%				
Children	3%				
Other	9%				
Publication type				93	0%
Journal article	33%				
Proceedings	55%				
Book chapter	3%				
Thesis	6%				
Other	2%				
Response scales				52	44%
5-point	89%				
6-point	4%				
7-point	8%				
Number of items for anthropomorphism				40	57%
2 items	3%				
3 items	5%				
5 items	93%				
Number of items for likeability				39	58%
4 items	3%				
5 items	90%				
6 items[a]	8%				

Note. Valid = Number of samples that reported the respective information. Missing = Percentage of samples failing to report the respective information.
[a]We suspect the studies by the research group claiming to have administered a sixth item (Foster et al., 2012; Giuliani et al., 2013) to be a reporting error because Bartneck, Kulić, and colleagues (2009) did not present a sixth item.

The nonlinear relationship suggested by the uncanny valley hypothesis (see Figure 1) could be modeled using higher-order polynomials of degree 3. To empirically determine the optimal number of higher-order terms, different meta-regression models were estimated and compared using the Bayesian information criterion (Schwarz, 1978). This suggested the inclusion of a linear term, a quadratic term, and a cubic term (see ESM 1). The respective meta-regression revealed a significant ($p < .05$) effect for anthropomorphism ($Q_m = 89.46$, $df = 3$, $p < .001$) that explained about 5% in the variance of likeability ratings between samples (see Table 2). These results were rather robust ($Q_m = 98.43$, $df = 3$, $p < .001$) and replicated after controlling for sample characteristics (i.e., mean age, share of women, publication year, country), robot characteristics (i.e., movement, communication), and methodological characteristics (i.e., presentation mode, risk of study bias). To study the effect in more detail, the likeability ratings predicted from this meta-regression model (including a 95% confidence interval) were plotted in Figure 4. Consistent with the assumption (a), these results confirmed more positive evaluations for more human-like robots overall. In accordance

© 2022 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY-NC-ND 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000486 - Thursday, March 28, 2024 3:48:59 PM - IP Address:44.204.94.166

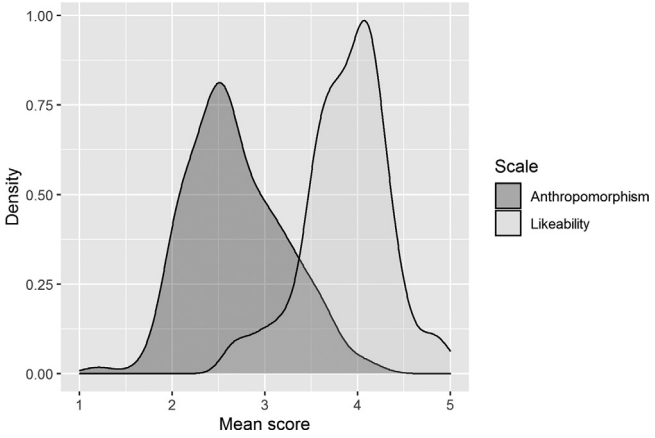


Figure 2. Average score distributions of the Godspeed anthropomorphism and likeability scales.

with the assumption (b), we also found evidence for a nonlinear effect. Although the effect approximated a sigmoid shape with a plateau in the region of the greatest anthropomorphism scores contained in the sample, we were unable to corroborate the hypothesized decline of likeability for highly realistic android robots as stated in assumption (c). Consequently, we were also unable to identify the rise of likeability at even higher scores of human likeness as expected in assumption (d). Again, these results were rather stable and replicated after controlling for various covariates (see Figure 4 and Table 2). The pooled association between anthropomorphism and likeability was also rather invariant toward various methodological choices and replicated after excluding outliers, children, or older samples and adopting robust meta-analytic models (see ESM 1).

Movement and Other Moderating Effects

In line with Mori's hypothesis (Mori et al., 2012), static robots were evaluated significantly less human-like as compared to moving robots (B = −0.35, 95% CI [−0.56, −0.14]). In contrast, the movement had no impact on likeability ratings (see Table E2 in ESM 1). Unexpectedly, communication had an opposite effect: For anthropomorphism, it was immaterial whether a robot was mute or communicated with the participants (B = 0.23, 95% CI [−0.07, 0.54]), whereas communicative robots were evaluated significantly ($p < .05$) more likable as compared to mute robots (B = −0.27, 95% CI [−0.47, −0.06]). To examine whether these effects also extended to the nonlinear association between anthropomorphism and likeability, we extended the previous meta-regression analyses and included respective interactions for the linear, quadratic, and cubic terms. However, inconsistent with the assumption (e), these interactions were not significant (see Table 2), thus, indicating that movement and communication did not moderate the predicted effects given in Figure 4. However, our database included only 19 results with static robots, while most of the robots exhibited some form of movement.

Discussion

Masahiro Mori's (1970) hypothetical graph on the uncanny valley has developed into a dominant influence on recent research into user perceptions of human-like robots. Complementing and extending insights gained from narrative reviews on the uncanny valley hypothesis (Kätsyri et al., 2015; Wang et al., 2015; Zhang et al., 2020), we presented

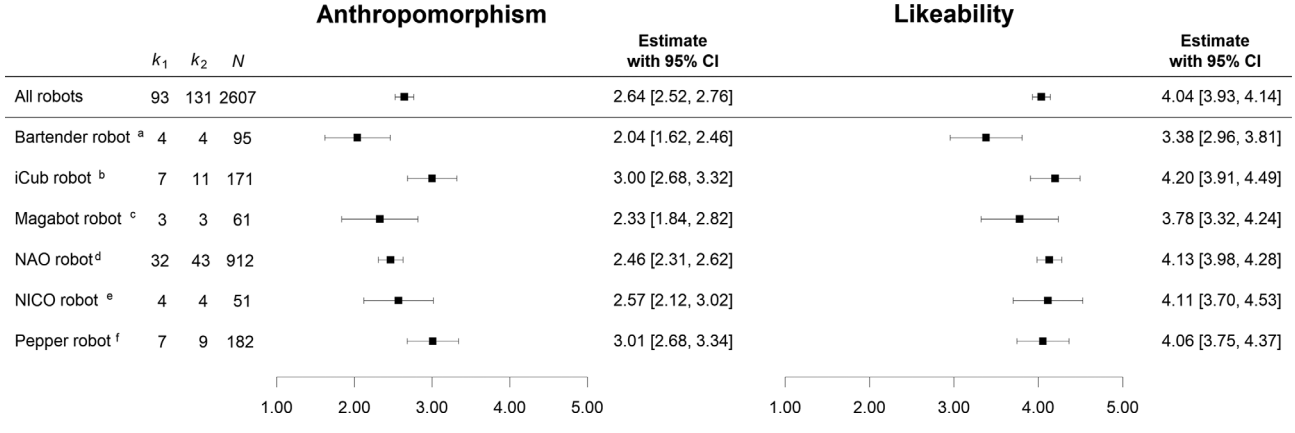


Figure 3. Forest plots for average anthropomorphism and likeability scores by robot model. k_1 = Number of samples, k_2 = number of ratings, N = total sample size. [a]Foster et al. (2012), Giuliani et al. (2013), Keizer et al. (2014); [b]Ghiglino et al. (2020), Lehmann et al. (2016), Mazzola et al. (2020), Willemse & Wykowska (2019); [c]Hoegen (2013), Lohse et al. (2013); [d]Barlas (2019), Cuijpers et al. (2011), Ham et al. (2015), van der Hout (2017), Lehmann et al. (2020), Mirnig, Stollnberger, Giuliani, et al. (2017), Mirnig, Stollnberger, Miksch et al. (2017), Rosenberg-Kima et al. (2020), Rosenthal-von der Pütten et al. (2017, 2018), Schneider (2019), Zanatto, Patacchiola, Goslin, & Cangelosi (2019), Zanatto, Patacchiola, Goslin, Thill, & Cangelosi (2020); [e]Churamani et al. (2017), Kerzel et al. (2020); [f]Iwashita & Katagami (2020), Rhim et al. (2019), Straßmann et al. (2020).

© 2022 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY-NC-ND 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000486 - Thursday, March 28, 2024 3:48:59 PM - IP Address:44.204.94.166

Table 2. Polynomial meta-regression tests for the Uncanny Valley Hypothesis

	Model 1	Model 2	Model 3	Model 4
Intercept	8.94*** (2.10)	8.72*** (2.18)	8.66* (3.53)	6.95*** (2.07)
Anthropomorphism				
1. Linear term	−6.19** (2.32)	−5.57** (2.41)	−5.93^{+} (3.75)	−3.53 (2.31)
2. Quadratic term	2.28** (0.85)	2.03* (0.88)	2.20 (1.30)	1.23 (0.84)
3. Cubic term	−0.25* (0.10)	−0.22* (0.11)	−0.24 (0.15)	−0.12 (0.10)
Control variables				
4. Average age[a]		0.00 (0.00)		
5. Share of women[b]		0.58* (0.26)		
Country[c]				
6. United Kingdom		−0.17 (0.16)		
7. Other country		−0.23* (0.09)		
8. Publication year[d]		0.00 (0.02)		
9. Movement[e]		−0.14^{+} (0.07)	7.23 (5.37)	
10. Communication[e]		−0.17* (0.08)		5.89 (8.80)
11. Interaction with real robot[e]		0.02 (0.09)		
12. Statistics reported[e]		−0.16 (0.11)		
13. Risk of study bias[f]		−0.06 (0.04)		
Moderating effects				
14. 1. × 8.			−10.60 (6.75)	
15. 2. × 8.			4.87^{+} (2.84)	
16. 3. × 8.			−0.72^{+} (0.40)	
17. 1. × 9.				−7.98 (9.67)
18. 2. × 9.				3.16 (3.50)
19. 3. × 9.				−0.39 (0.42)
Random effects (τ_s/τ_e)	0.39/0.08	0.35/0.04	0.40/0.04	0.37/0.04
I^2	96%	95%	96%	95%
R^2	5%	23%***	3%***	17%***

Note. Dependent variable are likeability ratings. Presented are meta-regression coefficients with standard errors in parentheses. τ_s/τ_e = Standard deviations of random effects for samples and evaluations; R^2 = Explained random variance. [a]Centered at 25 years, [b]Centered at .50, [c]Dummy coded with Germany as reference category, [d]Centered at year 2020, [e]0 = yes, 1 = no, [f]Centered at 4. ***$p < .001$; ** $p < .01$; * $p < .05$; $^{+}p < .10$.

the first quantitative, meta-analytical review of the main assumptions underlying the uncanny valley effect. We focused on the characteristic relationship between user assessments of human likeness (the *x*-axis) and likeability (the *y*-axis) that was proposed by Mori (1970, Figure 1), based on the Godspeed scales (Bartneck, Kulić, et al., 2009), a standard measure in the field (see Weiss & Bartneck, 2015). To this end, state-of-the-art meta-analytic methods that acknowledged dependencies between samples using a random-effects structure (see Cheung, 2019; Van den Noortgate et al., 2013) were used to study the nonlinear hypothesis with polynomial meta-regression analyses. From our quantitative assessment of the 93 independent samples that comprised our meta-analytic database, a main insight is the limited range of anthropomorphism and likeability scores in the examined primary studies (Figure 2). In the large majority of studies, the focal robot was experienced as being not quite human-like with means ranging below the scale's midpoint. Means above 3.5 on a 5-point scale were almost entirely missing. Likewise, and even more pronounced, the focal robots were experienced as highly likable on average in the primary studies. The large majority of studies reported mean likeability scores above the midpoint of the scale. The limited range of the primary study scores is highly relevant for our main meta-analytic aim, gathering quantitative evidence for or against the uncanny valley hypothesis. According to Mori (1970) and contemporary interpretations of his ideas (e.g., Diel & MacDorman, 2021; Wang et al., 2015), the characteristic drop in likeability is experienced at the higher end of the human likeness continuum. Based on the studies underlying our meta-analysis, this higher end of the human likeness continuum is unchartered territory.

We deduced several functional properties from the curvilinear explication of the uncanny valley hypothesis. Despite the identified limitations in scale range, likeability scores supported the first assumption (a) derived from Mori's uncanny valley hypothesis in that increasing human likeness was found to be associated with increased positive user responses within the spectrum of low to medium

© 2022 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY-NC-ND 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000486 - Thursday, March 28, 2024 3:48:59 PM - IP Address:44.204.94.166

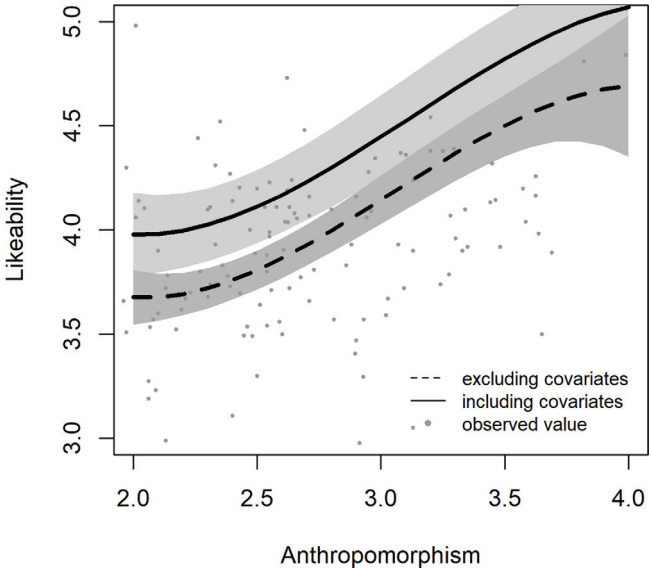


Figure 4. Predicted likeability ratings with 95% confidence intervals.

anthropomorphism. Important against the backdrop of the uncanny valley literature and in line with the assumption (b), our results also suggest a nonlinear effect, leading to a flattening of the likeability curve at about 75% of the anthropomorphism scale range (*x*-axis). However, because hardly any robots had been rated as highly human-like in the available primary studies, neither assumption (c) that such robots would lead to a pronounced drop in acceptance nor assumption (d) that near-to-perfect copies of humans at the end of the continuum would lead to an ultimate grow in acceptance could be evaluated. Mori's core proposition about the adverse effects of android robots can therefore neither be rejected nor confirmed at this stage.

We further examined several potential moderating variables. A comparison between static and moving robots was of particular relevance to the original uncanny valley hypothesis. Static robots were evaluated as less human-like than moving robots, but the movement had no impact on likeability ratings. Importantly, the linear, quadratic, and cubic associations between human likeness and likeability did not differ significantly between statically presented robots and such that were moving. Assumption (e), based on Mori's description of a potentially intensifying role of robotic motion, therefore must be rejected in view of the current data.

Limitations and Implications

As outlined above, our quantitative test of the uncanny valley hypothesis is preliminary, as primary studies that captured high degrees of human likeness were missing. The low human likeness scores observed could be a function of several factors. First, the robotic platforms examined in the primary studies do not stipulate high human likeness (e.g., NAO and similar designs, see ESM 1). Second, participants naïve to robotics may use expectations derived from science-fiction as a point of comparison (Appel et al., 2016; Mara & Appel, 2015). Due to the fact that the state of today's technological advancement rarely matches sci-fi worlds, robots examined in human-robot interaction research have to fall short compared to fictional robots. The original movie Blade Runner (Scott, 1982), for example, showed a world in the year 2019 in which humans and human-like robotic replicants mingled. Participants with high technological knowledge or even a study emphasis in computer science, in turn, may be aware of technological glitches or wizard-of-oz simulated interactions.

We deliberately restricted our study pool to primary studies that reported data on the Godspeed Scales (Bartneck, Kulić, et al., 2009) to achieve high comparability and to prevent an influx of data with low reliability or validity, which has been described as a substantial problem in the field (Wang et al., 2015). The Godspeed Scales are in particular widespread use, constituting one of the standard measures in the field. Despite their popularity, it should not be dismissed that the Godspeed Scales themselves have also faced some criticism in the past (Carpinella et al., 2017; Ho & MacDorman, 2010). For example, an exploratory factor analysis conducted by Carpinella and colleagues (2017) indicated low eigenvalues and low reliabilities for some of the five Godspeed components. However, this was mainly true for the animacy and safety scales, but not for anthropomorphism and likeability. Consistent with this and in support of our decision to use the Godspeed Sales, our database showed high reliabilities for both the anthropomorphism scale and the likeability scale. That said, future meta-analyses could apply more liberal inclusion criteria. Promising alternative measures include the scales by Ho and MacDorman (2010, 2017), which were developed specifically for research on the uncanny valley hypothesis, or the Robotic Social Attributes Scale (Carpinella et al., 2017), which assesses warmth and competence as components of social perception and discomfort as a potential measure for the uncanny experience.

We further restricted our meta-analysis to genuine implementations of robotic systems. Studies that relied on verbal descriptions, drawings of robots, or morphed pictures (e.g., Lischetzke et al., 2017; MacDorman & Ishiguro, 2006) were excluded. Whereas these stimuli could arguably increase human likeness (e.g., morphs between robots and humans, Lischetzke et al., 2017), such stimuli have been criticized for lacking external validity, for example, morphs may show ghosting artifacts by the computer graphics software (Kätsyri et al., 2019).

Several measures were taken to secure a standard of sufficient data quality in the primary study pool and, therefore,

© 2022 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY-NC-ND 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000486 - Thursday, March 28, 2024 3:48:59 PM - IP Address:44.204.94.166

our meta-analysis as a summary of the quantitative results. This includes the restriction to the experience of genuine technical implementations and the Godspeed Scales as operationalizations of the key variables. We further implemented a risk of study bias assessment (Nudelman & Otto, 2020) and controlled our meta-analytic results for the respective scores. These scores revealed remarkable weaknesses regarding design or reporting. We need to acknowledge these shortcomings of the primary study data, and we emphasize two implications for human-robot interaction research:

First, our review of studies revealed that a substantial number of publications failed to report basic information on the sample and descriptive results. Authors of quantitative results sections should make sure to report (subgroup-) sample sizes and results on variance (e.g., the standard deviation) along with mean values (or any other measure of central tendency). Zero-order correlations and raw descriptive statistics are particularly helpful for (meta-analytic) summaries and comparisons within a field of research. Second, sample sizes were remarkably small, $Mdn(N) = 21$, from a general psychological perspective. They arguably reflect the studied topic in human-robot interaction research for which the technological requirements complicate or impede larger sample sizes. Nevertheless, minimal sample size recommendations should be adhered to (Simmons et al., 2011). Note that 20 participants per cell, for example, is insufficient to "detect in a representative sample that men are heavier than women" (Simmons et al., 2018, p. 256). The problem of low sample size is even more serious for complex between-subjects designs (e.g., a focal moderation effect based on a 2×2 experimental design). The authors of several other recent meta-analyses and reviews in the field of human-robot interaction also identified similar problems in data reporting and statistical power of primary studies and made similar recommendations to the interdisciplinary research community (Leichtmann & Nitsch, 2020; Oliveira et al., 2021; Stower et al., 2021). We are therefore optimistic that future empirical work will benefit from the lessons learned and, through larger sample sizes and greater transparency, will make important contributions to our understanding of user responses to robots.

Conclusion

The uncanny valley hypothesis is a major perspective to explaining and predicting negative responses to humanoid and android robots. The available research covers user experiences of low to moderate human likeness, whereas robots with high human likeness are largely unchartered territory. Within these low to moderate levels of human likeness, our findings follow the assumptions derived from the uncanny valley hypothesis insofar as likeability ratings initially increase but then level off to a plateau as a result of a nonlinear function. Movement appears to be no factor that intensifies the characteristic nonlinear association between human likeness and likeability.

Electronic Supplementary Material

The electronic supplementary material is available with the online version of the article at https://doi.org/10.1027/2151-2604/a000486

ESM 1. Coded variables and data, diagrams, Meta-Analyses of Godspeed Scale Scores by Robot, and further analyses. Table E2: Meta-Analyses of Godspeed Scale Scores by Robot Model.

References

References marked with * were included in the meta-analysis.

Appel, M., Izydorczyk, D., Weber, S., Mara, M., & Lischetzke, T. (2020). The uncanny of mind in a machine: Humanoid robots as tools, agents, and experiencers. *Computers in Human Behavior, 102*, 274–286. https://doi.org/10.1016/j.chb.2019.07.031

Appel, M., Krause, S., Gleich, U., & Mara, M. (2016). Meaning through fiction: Science fiction and innovative technologies. *Psychology of Aesthetics, Creativity, and the Arts, 10*, 472–480. https://doi.org/10.1037/aca0000052

*Avelino, J., Correia, F., Catarino, J., Ribeiro, P., Moreno, P., Bernardino, A., & Paiva, A. (2018). The power of a hand-shake in human-robot interactions. In *Proceedings of the 2018 International Conference on Intelligent Robots and Systems* (pp. 1864–1869). IEEE. https://doi.org/10.1109/IROS.2018.8593980

*Barlas, Z. (2019). When robots tell you what to do: Sense of agency in human-and robot-guided actions. *Consciousness and Cognition, 75*, Article 102819. https://doi.org/10.1016/j.concog.2019.102819

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

Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2009). My robotic doppelgänger – A critical look at the uncanny valley. In *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 269–276). IEEE. https://doi.org/10.1109/ROMAN.2009.5326351

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

*Busch, B., Cotugno, G., Khoramshahi, M., Skaltsas, G., Turchi, D., Urbano, L., Wächter, M., Zhou, Y., Asfour, T., Deacon, G., Russell, D., & Billard, A. (2019). Evaluation of an industrial robotic assistant in an ecological environment. In *Proceedings of the 28th IEEE International Conference on Robot and Human Interactive Communication* (pp. 1–8). IEEE. https://doi.org/10.1109/RO-MAN46459.2019.8956399

© 2022 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY-NC-ND 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000486 - Thursday, March 28, 2024 3:48:59 PM - IP Address:44.204.94.166

Čapek, K. (1920/2001). *R.U.R. (Rossum's Universal Robots)*. Dover.

Carpinella, C. M., Wyman, A. B., Perez, M. A., & Stroessner, S. J. (2017). The Robotic Social Attributes Scale (RoSAS): Development and validation. In *Proceedings of the International Conference on Human-Robot Interaction* (pp. 254–262). IEEE. https://doi.org/10.1145/2909824.3020208

Cave, S., Dihal, K., & Dillon, S. (2020). Introduction: Imagining AI. In S. Cave, K. Dihal, & S. Dillon (Eds.), *AI Narratives: A history of imaginative thinking about intelligent machines* (pp. 1–21). . Oxford University Press. https://doi.org/10.1093/oso/9780198846666.001.0001

Cheung, M. W. L. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology Review, 29*(4), 387–396. https://doi.org/10.1007/s11065-019-09415-6

*Churamani, N., Anton, P., Brügger, M., Fließwasser, E., Hummel, T., Mayer, J., Mustafa, W., Ng, H. G., Nguyen, T. L. C., Nguyen, Q., Soll, M., Springenberg, S., Griffiths, S., Heinrich, S., Navarro-Guerrero, N., Strahl, E., Twiefel, J., Weber, C., & Wermter, S. (2017). The impact of personalisation on human-robot interaction in learning scenarios. In B. Wrede, Y. Nagai, T. Komatsu, M. Hanheide, & L. Natale (Eds.), *Proceedings of the 5th International Conference on Human Agent Interaction* (pp. 171–180). Association for Computing Machinery. https://doi.org/10.1145/3125739.3125756

*Cuijpers, R. H., Bruna, M. T., Ham, J. R., & Torta, E. (2011). Attitude towards robots depends on interaction but not on anticipatory behaviour. In B. Mutlu, C. Bartneck, J. Ham, V. Evers, & T. Kanda (Eds.), *Proceedings of the 2011 International Conference on Social Robotics* (pp. 163–172). . Springer. https://doi.org/10.1007/978-3-642-25504-5_17

Diel, A., & MacDorman, K. F. (2021). Creepy cats and strange high houses: Support for configural processing in testing predictions of nine uncanny valley theories. *Journal of Vision, 21*(4), 1–20. https://doi.org/10.1167/jov.21.4.1

Döring, N., Mohseni, M. R., & Walter, R. (2020). Design, use, and effects of sex dolls and sex robots: Scoping review. *Journal of Medical Internet Research, 22*(7), Article e18551. https://doi.org/10.2196/18551

*Foster, M. E., Gaschler, A., Giuliani, M., Isard, A., Pateraki, M., & Petrick, R. P. (2012). Two people walk into a bar: Dynamic multi-party social interaction with a robot agent. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction* (pp. 3–10). ACM. https://doi.org/10.1145/2388676.2388680

*Fu, C., Yoshikawa, Y., Iio, T., & Ishiguro, H. (2021). Sharing experiences to help a robot present its mind and sociability. *International Journal of Social Robotics, 13*, 341–352. https://doi.org/10.1007/s12369-020-00643-y

*Ghiglino, D., De Tommaso, D., Willemse, C., Marchesi, S., & Wykowska, A. (2020). Can I get your (robot) attention? Human sensitivity to subtle hints of human-likeness in a humanoid robot's behavior. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd Annual Virtual Meeting of the Cognitive Science Society* (pp. 952–958). Cognitive Science Society. https://doi.org/10.31234/osf.io/kfy4g

*Giuliani, M., Petrick, R. P., Foster, M. E., Gaschler, A., Isard, A., Pateraki, M., & Sigalas, M. (2013). Comparing task-based and socially intelligent behaviour in a robot bartender. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction* (pp. 263–270). ACM. https://doi.org/10.1145/2522848.2522869

Gnambs, T. (2019). Attitudes towards emergent autonomous robots in Austria and Germany. *Elektrotechnik und Informationstechnik, 136*, 296–300. https://doi.org/10.1007/s00502-019-00742-3

Gnambs, T., & Appel, M. (2019). Are robots becoming unpopular? Changes in attitudes towards autonomous robotic systems in Europe. *Computers in Human Behavior, 93*, 53–61. https://doi.org/10.1016/j.chb.2018.11.045

*Ham, J., Cuijpers, R. H., & Cabibihan, J. J. (2015). Combining robotic persuasive strategies: The persuasive power of a storytelling robot that uses gazing and gestures. *International Journal of Social Robotics, 7*(4), 479–487. https://doi.org/10.1007/s12369-015-0280-4

*Haring, K. S., Silvera-Tawil, D., Takahashi, T., Velonaki, M., & Watanabe, K. (2015). Perception of a humanoid robot: a cross-cultural comparison. In *Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 821–826). IEEE. https://doi.org/10.1109/ROMAN.2015.7333613

*Haring, K. S., Silvera-Tawil, D., Takahashi, T., Watanabe, K., & Velonaki, M. (2016). How people perceive different robot types: A direct comparison of an android, humanoid, and non-biomimetic robot. In *Proceedings of the 8th International Conference on Knowledge and Smart Technology* (pp. 265–270). IEEE. https://doi.org/10.1109/KST.2016.7440504

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*, 39–65. https://doi.org/10.1002/jrsm.5

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

Ho, C. C., & MacDorman, K. F. (2017). Measuring the uncanny valley effect. *International Journal of Social Robotics, 9*(1), 129–139. https://doi.org/10.1007/s12369-016-0380-9

*Hoegen, R. (2013). *The influence of a robot's voice on proxemics in human-robot interaction* [Unpublished manuscript]. University of Twente.

Hoffmann, M., & Pfeifer, R. (2018). Robots as powerful allies for the study of embodied cognition from the bottom up. In A. Newen, L. de Bruin, & S. Gallagher (Eds.), *The Oxford handbook of 4E cognition* (pp. 841–862). . Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198735410.013.45

Ishiguro, H. (2016). Android science. In M. Kasaki, H. Ishiguro, M. Asada, M. Osaka, & T. Fujikado (Eds.), *Cognitive neuroscience robotics A: Synthetic approaches to human understanding* (pp. 193–234). Springer. https://doi.org/10.1007/978-4-431-54595-8

*Iwashita, M., & Katagami, D. (2020). Psychological effects of compliment expressions by communication robots on humans. In *Proceedings of the 2020 International Joint Conference on Neural Networks* (pp. 1–8). IEEE. https://doi.org/10.1109/IJCNN48605.2020.9206898

*Johanson, D. L., Ahn, H. S., Lim, J., Lee, C., Sebaratnam, G., MacDonald, B. A., & Broadbent, E. (2020). Use of humor by a healthcare robot positively affects user perceptions and behavior. *Technology, Mind, and Behavior, 1*(2), 1–33. https://doi.org/10.1037/tmb0000021

Johnson, D. O., Cuijpers, R. H., Pollmann, K., & van de Ven, A. A. (2016). Exploring the entertainment value of playing games with a humanoid robot. *International Journal of Social Robotics, 8*(2), 247–269.

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

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

© 2022 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY-NC-ND 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000486 - Thursday, March 28, 2024 3:48:59 PM - IP Address:44.204.94.166

eeriness. *Frontiers in Psychology, 6*, Article 390. https://doi.org/10.3389/fpsyg.2015.00390

*Keizer, S., Foster, M. E., Gaschler, A., Giuliani, M., Isard, A., & Lemon, O. (2014). Handling uncertain input in multi-user human-robot interaction. In *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication* (pp. 312–317). IEEE. https://doi.org/10.1109/ROMAN.2014.6926271

*Kerzel, M., Pekarek-Rosin, T., Strahl, E., Heinrich, S., & Wermter, S. (2020). Teaching NICO how to grasp: An empirical study on crossmodal social interaction as a key factor for robots learning from humans. *Frontiers in Neurorobotics, 14*, Article 28. https://doi.org/10.3389/fnbot.2020.00028

Kim, B., Bruce, M., Brown, L., de Visser, E., & Phillips, E. (2020). A comprehensive approach to validating the uncanny valley using the Anthropomorphic RoBOT (ABOT) database. In *Proceedings of 2020 Systems and Information Engineering Design Symposium (SIEDS)* (pp. 1–6). IEEE. https://doi.org/10.1109/SIEDS49339.2020.9106675

*Kühnlenz, B. (2013). *Alignment strategies for information retrieval in prosocial human-robot interaction* [Unpublished doctoral dissertation]. Technical University Munich.

*Kühnlenz, B., Sosnowski, S., Buß, M., Wollherr, D., Kühnlenz, K., & Buss, M. (2013). Increasing helpfulness towards a robot by emotional adaption to the user. *International Journal of Social Robotics, 5*(4), 457–476. https://doi.org/10.1007/s12369-013-0182-2

*Lehmann, H., Rojik, A., & Hoffmann, M. (2020, September). *Should a small robot have a small personal space? Investigating personal spatial zones and proxemic behavior in human-robot interaction*. Paper presented at the CognitIve RobotiCs for intEraction (CIRCE) Workshop at the 2020 IEEE International Conference on Robot and Human Interactive Communication. https://arxiv.org/abs/2009.01818

*Lehmann, H., Roncone, A., Pattacini, U., & Metta, G. (2016). Physiologically inspired blinking behavior for a humanoid robot. In A. Agah, J. J. Cabibihan, A. Howard, M. Salichs, & H. He (Eds.), *Proceedings of the 2016 International Conference on Social Robotics* (pp. 83–93). . Springer. https://doi.org/10.1007/978-3-319-47437-3_9

Leichtmann, B., & Nitsch, V. (2020). How much distance do humans keep toward robots? Literature review, meta-analysis, and theoretical considerations on personal space in human-robot interaction. *Journal of Environmental Psychology, 68*, 101386. https://doi.org/10.1016/j.jenvp.2019.101386

Lischetzke, T., Izydorczyk, D., Hüller, C., & Appel, M. (2017). The topography of the uncanny valley and individuals' need for structure: A nonlinear mixed effects analysis. *Journal of Research in Personality, 68*, 96–113. https://doi.org/10.1016/j.jrp.2017.02.001

*Löffler, D., Hurtienne, J., & Nord, I. (2019). Blessing robot blessU2: A discursive design study to understand the implications of social robots in religious contexts. *International Journal of Social Robotics*. Advance online publication. doi: https://doi.org/10.1007/s12369-019-00558-3

*Lohse, M., van Berkel, N., Van Dijk, E. M., Joosse, M. P., Karreman, D. E., & Evers, V. (2013). The influence of approach speed and functional noise on users' perception of a robot. In *Proceedings of the 2013 International Conference on Intelligent Robots and Systems* (pp. 1670–1675). IEEE. https://doi.org/10.1109/IROS.2013.6696573

*Lugrin, B., Dippold, J., & Bergmann, K. (2018). Social robots as a means of integration? An explorative acceptance study considering gender and non-verbal behaviour. In *Proceedings of the 2018 International Conference on Intelligent Robots and Systems* (pp. 2026–2032). IEEE. https://doi.org/10.1109/IROS.2018.8593818

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

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

Mara, M., Schreibelmayr, S., & Berger, F. (2020). Hearing a nose? User expectations of robot appearance induced by different robot voices. In *Proceedings of the Companion of the 2020 International Conference on Human-Robot Interaction* (pp. 355–356). ACM/IEEE. https://doi.org/10.1145/3371382.3378285

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

*Mazzola, C., Aroyo, A. M., Rea, F., & Sciutti, A. (2020). Interacting with a social robot affects visual perception of space. In *Proceedings of the 2020 ACM/IEEE International Conference on Human Robot Interaction* (pp. 549–557). ACM. https://doi.org/10.1145/3319502.3374819

*Meghdari, A., Shariati, A., Alemi, M., Vossoughi, G. R., Eydi, A., Ahmadi, E., Mozafari, B., Nobaveh, A. A., & Tahami, R. (2018). Arash: A social robot buddy to support children with cancer in a hospital environment. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 232*(6), 605–618. https://doi.org/10.1177/0954411918777520

*Mirnig, N., Stollnberger, G., Giuliani, M., & Tscheligi, M. (2017). Elements of humor: How humans perceive verbal and non-verbal aspects of humorous robot behavior. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 211–212). Association for Computing Machinery. https://doi.org/10.1145/3029798.3038337

*Mirnig, N., Stollnberger, G., Miksch, M., Stadler, S., Giuliani, M., & Tscheligi, M. (2017). To err is robot: How humans assess and act toward an erroneous social robot. *Frontiers in Robotics and AI, 4*, Article 21. https://doi.org/10.3389/frobt.2017.00021

*Moon, A., Parker, C. A., Croft, E. A., & Van der Loos, H. M. (2013). Design and impact of hesitation gestures during human-robot resource conflicts. *Journal of Human-Robot Interaction, 2*(3), 18–40. https://doi.org/10.5898/JHRI.2.3.Moon

Mori, M. (1970). Bukimi no tani [The uncanny valley]. *Energy, 7*, 33–35.

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

*Müller, S. L., Schröder, S., Jeschke, S., & Richert, A. (2017). Design of a robotic workmate. In V. Duffy (Ed.), *International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management* (pp. 447–456). Springer. https://doi.org/10.1007/978-3-319-58463-8_37

Nudelman, G., & Otto, K. (2020). The development of a new generic risk-of-bias measure for systematic reviews of surveys. *Methodology, 16*, 278–298. https://doi.org/10.5964/meth.4329

Ogawa, K., Nishio, S., Koda, K., Taura, K., Minato, T., Ishii, C. T., & Ishiguro, H. (2011). Telenoid: Tele-presence android for communication. In *Proceedings of the SIGGRAPH 2011 Emerging Technologies* (pp. 1). ACM. https://doi.org/10.1145/2048259.2048274

Oliveira, R., Arriaga, P., Santos, F. P., Mascarenhas, S., & Paiva, A. (2021). Towards prosocial design: A scoping review of the use of robots and virtual agents to trigger prosocial behaviour. *Computers in Human Behavior, 114*, Article 106547. https://doi.org/10.1016/j.chb.2020.106547

*Paetzel, M., Perugia, G., & Castellano, G. (2020). The persistence of first impressions: The effect of repeated interactions on the

© 2022 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY-NC-ND 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000486 - Thursday, March 28, 2024 3:48:59 PM - IP Address:44.204.94.166

perception of a social robot. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 73–82). ACM. https://doi.org/10.1145/3319502.3374786

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal, 372*, Article n71. https://doi.org/10.1136/bmj.n71

Pandey, A. K., & Gelin, R. (2018). A mass-produced sociable humanoid robot: Pepper: The first machine of its kind. *IEEE Robotics & Automation Magazine, 25*(3), 40–48. https://doi.org/10.1109/MRA.2018.2833157

*Petrak, B., Weitz, K., Aslan, I., & André, E. (2019). Let me show you your new home: Studying the effect of proxemic-awareness of robots on users' first impressions. In *Proceedings of the 28th IEEE International Conference on Robot and Human Interactive Communication* (pp. 1–7). IEEE. https://doi.org/10.1109/RO-MAN46459.2019.8956463

Pick, J. L., Nakagawa, S., & Noble, D. W. (2019). Reproducible, flexible and high-throughput data extraction from primary literature: The *metaDigitise* R package. *Methods in Ecology and Evolution, 10*, 426–431. https://doi.org/10.1111/2041-210X.13118

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*(3), 271–277. https://doi.org/10.1016/j.cognition.2013.11.001

R Core Team. (2020). *R: A language and environment for statistical computing* (Version 4.0.3) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org

*Rhim, J., Cheung, A., Pham, D., Bae, S., Zhang, Z., Townsend, T., & Lim, A. (2019). Investigating positive psychology principles in affective robotics. In *Proceedings of the 8th International Conference on Affective Computing and Intelligent Interaction* (pp. 1–7). IEEE. https://doi.org/10.1109/ACII.2019.8925475

*Rosenberg-Kima, R. B., Koren, Y., & Gordon, G. (2020). Robot-supported collaborative learning (RSCL): Social robots as teaching assistants for higher education small group facilitation. *Frontiers in Robotics and AI, 6*, Article 148. https://doi.org/10.3389/frobt.2019.00148

*Rosenthal-von der Pütten, A. M., Bock, N., & Brockmann, K. (2017). Not your cup of tea? How interacting with a robot can increase perceived self-efficacy in HRI and evaluation. In *Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 483–492). IEEE. https://doi.org/10.1145/2909824.3020251

*Rosenthal-von der Pütten, A. M., Krämer, N. C., & Herrmann, J. (2018). The effects of humanlike and robot-specific affective nonverbal behavior on perception, emotion, and behavior. *International Journal of Social Robotics, 10*(5), 569–582. https://doi.org/10.1007/s12369-018-0466-7

*Ruijten, P. A., & Cuijpers, R. H. (2018). If drones could see: Investigating evaluations of a drone with eyes. In S. S. Ge, J.-J. Cabibihan, M. A. Salichs, E. Broadbent, H. He, A. R. Wagner, & Á. Castro-González (Eds.), *Proceedings of the 10th International Conference on Social Robotics* (pp. 65–74). Springer. https://doi.org/10.1007/978-3-030-05204-1_7

*Schneider, S. (2019). *Socially assistive robots for exercise scenarios* [Unpublished dissertation]. Bielefeld University. https://doi.org/10.4119/unibi/2934006

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Scott, R. (. Director). (1982). *Blade runner* [movie]. Warner Bros.

*Shariati, A., Shahab, M., Meghdari, A., Nobaveh, A. A., Rafatnejad, R., & Mozafari, B. (2018). Virtual reality social robot platform: A case study on Arash social robot. In S. S. Ge, J.-J. Cabibihan, M. A. Salichs, E. Broadbent, H. He, A. R. Wagner, & Á. Castro-González (Eds.), *Proceedings of the 10th International Conference on Social Robotics* (pp. 551–560). Springer. https://doi.org/10.1007/978-3-030-05204-1_54

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359–1366. https://doi.org/10.1177/0956797611417632

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2018). False-positive citations. *Perspectives on Psychological Science, 13*(2), 255–259. https://doi.org/10.1177/1745691617698146

Stein, J. P., & Ohler, P. (2017). Venturing into the uncanny valley of mind – The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition, 160*, 43–50. https://doi.org/10.1016/j.cognition.2016.12.010

Stower, R., Calvo-Barajas, N., Castellano, G., & Kappas, A. (2021). A meta-analysis on children's trust in social robots. *International Journal of Social Robotics*. Advance online publication. https://doi.org/10.1007/s12369-020-00736-8

*Straßmann, C., Grewe, A., Kowalczyk, C., Arntz, A., & Eimler, S. C. (2020). Moral robots? How uncertainty and presence affect humans' moral decision making. In C. Stephanidis & M. Antona (Eds.), *Proceedings of the 2020 International Conference on Human-Computer Interaction* (pp. 488–495). Springer. https://doi.org/10.1007/978-3-030-50726-8_64

*Syrdal, D. S., Dautenhahn, K., Koay, K. L., Walters, M. L., & Ho, W. C. (2013). Sharing spaces, sharing lives – the impact of robot mobility on user perception of a home companion robot. In G. Herrmann, M. J. Pearson, A. Lenz, P. Bremner, A. Spiers, & U. Leonards (Eds.), *Proceedings of the 2013 International Conference on Social Robotics* (pp. 321–330) . Springer. https://doi.org/10.1007/978-3-319-02675-6_32

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

*Trovato, G., Ramos, J. G., Azevedo, H., Moroni, A., Magossi, S., Ishii, H., Simmons, R., & Takanishi, A. (2015a). Designing a receptionist robot: Effect of voice and appearance on anthropomorphism. In *Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 235–240). IEEE. https://doi.org/10.1109/ROMAN.2015.7333573

*Trovato, G., Ramos, J. G., Azevedo, H., Moroni, A., Magossi, S., Ishii, H., Simmons, R., & Takanishi, A. (2015b). "Olá, my name is Ana": A study on Brazilians interacting with a receptionist robot. In *Proceedings for the 2015 International Conference on Advanced Robotics* (pp. 66–71). IEEE. https://doi.org/10.1109/ICAR.2015.7251435

*Ueno, A., Hlaváč, V., Mizuuchi, I., & Hoffmann, M. (2020). Touching a human or a robot? Investigating human-likeness of a soft warm artificial hand. In *Proceedings of the 29th IEEE International Conference on Robot and Human Interactive Communication* (pp. 14–20). IEEE. https://doi.org/10.1109/RO-MAN47096.2020.9223523

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods, 45*(2), 576–594. https://doi.org/10.3758/s13428-012-0261-6

*Van der Hout, V. M. (2017). *The touch of a robotic friend* [Unpublished master's thesis]. University of Twente. http://purl.utwente.nl/essays/73221

© 2022 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY-NC-ND 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0)

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1–48. https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1*(2), 112–125. https://doi.org/10.1002/jrsm.11

Voracek, M., Kossmeier, M., & Tran, U. S. (2019). Which data to meta-analyze, and how? A specification-curve and multiverse-analysis approach to meta-analysis. *Zeitschrift für Psychologie, 227*(1), 64–82. https://doi.org/10.1027/2151-604/a000357

Voskuhl, A. (2013). *Androids in the enlightenment: Mechanics, artisans, and cultures of the self*. University of Chicago Press.

Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology, 19*(4), 393–407. https://doi.org/10.1037/gpr0000056

Weiss, A., & Bartneck, C. (2015). Meta analysis of the usage of the Godspeed Questionnaire series. In *Proceedings of the 2015 International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 381–388). IEEE. https://doi.org/10.1109/ROMAN.2015.7333568

*Wieser, I., Toprak, S., Grenzing, A., Hinz, T., Auddy, S., Karaoğuz, E. C., Chandran, A., Remmels, M., Shinawi, A. E., Josifovski, J., Vankadara, L. C., Wahab, F. U., Bahnemiri, A. M., Sahu, D., Heinrich, S., Navarro-Guerrero, N., Strahl, E., Twiefel, J., & Wermter, S. (2016). A robotic home assistant with memory aid functionality. In G. Friedrich, M. Helmert, & F. Wotawa (Eds.), *Joint German/Austrian Conference on Artificial Intelligence* (pp. 102–115). Springer. https://doi.org/10.1007/978-3-319-46073-4_8

*Willemse, C., & Wykowska, A. (2019). In natural interaction with embodied robots, we prefer it when they follow our gaze: A gaze-contingent mobile eyetracking study. *Philosophical Transactions of the Royal Society B, 374*(1771), Article 20180036. https://doi.org/10.1098/rstb.2018.0036

Yoshikawa, M., Matsumoto, Y., Sumitani, M., & Ishiguro, H. (2011). Development of an android robot for psychological support in medical and welfare fields. In *Proceedings of the 2011 International Conference on Robotics and Biomimetics* (pp. 2378–2383). IEEE. https://doi.org/10.1109/ROBIO.2011.6181654

*Zanatto, D., Patacchiola, M., Goslin, J., & Cangelosi, A. (2019). Investigating cooperation with robotic peers. *PLoS One, 14*(11), Article e0225028. https://doi.org/10.1371/journal.pone.0225028

*Zanatto, D., Patacchiola, M., Goslin, J., Thill, S., & Cangelosi, A. (2020). Do humans imitate robots? An investigation of strategic social learning in human-robot interaction. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 449–457). Association for Computing Machinery. https://doi.org/10.1145/3319502.3374776

Zhang, J., Li, S., Zhang, J. Y., Du, F., Qi, Y., & Liu, X. (2020). A literature review of the research on the uncanny valley. In P.-L. P. Rau (Ed.), *Cross-cultural design. User experience of products, services, and intelligent environments* (pp. 255–268). Springer.

History
Received February 24, 2021
Revision received November 15, 2021
Accepted November 18, 2021
Published online February 2, 2022

Acknowledgments
The authors are grateful to the research assistants Christine Busch, Lina Curth, Laura Moradbakhti, Simon Schreibelmayr, and Sandra Siedl, who contributed to the coding of the primary studies included in this meta-analysis.

Conflict of Interest
We have no conflicts of interest to disclose.

Open Data
The meta-analysis was not preregistered. The data, material, computer code, and analysis results are publicly available at the Open Science Framework (OSF; https://osf.io/t9rdk).

ORCID
Martina Mara
https://orcid.org/0000-0003-3447-0556

Timo Gnambs
https://orcid.org/0000-0002-6984-1276

Martina Mara
LIT Robopsychology Lab
Johannes Kepler University Linz
Altenberger Straße 69
4040 Linz
Austria
martina.mara@jku.at

https://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000486 - Thursday, March 28, 2024 3:48:59 PM - IP Address:44.204.94.166

© 2022 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY-NC-ND 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0)