Improving Evaluations of Advanced Robots by Depicting Them in Harmful Situations

Andrea Grundke, Jan-Philipp Stein, and Markus Appel

Psychology of Communication and New Media

Julius-Maximilians-Universität Würzburg

Manuscript accepted for publication in the journal

Computers in Human Behavior (2022/11/14)

Corresponding author: Andrea Grundke

Oswald-Külpe-Weg 82, 97074 Würzburg (Germany).

E-Mail: andrea.grundke@uni-wuerzburg.de

Author Note: Data, materials, and supplements are publicly available under https://doi.org//10.17605/OSF.IO/G8BNW

Declaration of Interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding: Andrea Grundke was supported by a PhD-Scholarship from the Hanns-Seidel-Stiftung.

Highlights

- Robots with mind evoked aversion in prior studies (uncanny valley of mind)
- Two online experiments were conducted, testing human empathy as a countermeasure
- A robot shown in a harmful situation elicited higher empathy
- An indirect effect of the situation on likeability, via empathy, was observed
- A negative residual effect of showing the robot in a harmful situation emerged

Abstract

Equipping robots with sophisticated mental abilities can result in reduced likeability (*uncanny valley of mind*). Other work shows that exposing robots to harm increases empathy and likeability. Connecting both lines of research, we hypothesized that eliciting empathy could mitigate or even reverse the negative response to robots with mind. In two online experiments, we manipulated the attributes of a robot (with or without mind) and presented the robot in situations in which it was either exposed to harm or not. Perceived empathy for the robot and robot likeability served as dependent variables. Experiment 1 (N=559) used text vignettes to manipulate robot mind and a video that involved either physical harm or no harm to the machine. In a second experiment (N=396), both experimental factors were manipulated via the shown video. Across both experiments, we observed a significant indirect effect of presenting the robot in a harmful situation on likeability, with empathy serving as a mediating variable. Moreover, a residual negative influence of showing the robot in a harmful situation was detected. We conclude that the uncanny valley of mind observed in our studies could be based on the robot's human-like imperfection, rather than descriptions of its supposed mind.

Keywords: uncanny valley; mind perception; empathy; human-robot interaction; user acceptance

1. Introduction

Due to rapid technological advancements, robotic technology has become much more complex, diverse, and visible in recent years (International Federation of Robotics, 2021; Yang et al., 2020). At the same time, people's attitudes towards robots are not only ambivalent (Brondi et al., 2021; Stapels & Eyssel, 2022), empirical data suggest that attitudes towards robots have also become more negative in some parts of the world over the last years (Gnambs & Appel, 2019). As such, scientists from different disciplines are called upon to provide theory and empirical insight as to why people come to like or dislike certain robotic inventions. Offering a key piece of evidence in this regard, research indicates that perceiving mind (in terms of experience and agency) in a robotic machine may render it eerie and unlikeable (*uncanny valley of mind;* Appel et al., 2020; Dang & Liu, 2021; Gray & Wegner, 2012; Stein & Ohler, 2017). Yet, as novel robots are equipped with increasingly complex mental abilities to make them more social or useful (Bryndin, 2020; Hildt, 2019; Laird et al., 2017), it seems particularly worthwhile to find new ways of alleviating people's aversion to machines with sophisticated mental capacities.

One possible approach to this—and the focus of our work—is to evoke empathy for the robot. According to prior theory and research, presenting robots in situations in which they are harmed tends to elicit empathy and, in turn, more positive evaluations by observers (e.g., Cameron et al., 2021; Gonsior et al., 2011; Rosenthal-von der Pütten et al., 2013). This effect should be enhanced for robots with mind, as people's ability to empathize is intrinsically linked with perceptions of mental processes in others (Singer & Lamm, 2009). Therefore, the current project investigates whether depicting a robot that is physically harmed (vs. no harm) mitigates or even reverses the negative response to the robot in case it is described as possessing advanced mental abilities. Also, by using audiovisual stimuli, we pursue a more immersive manipulation than the text-based approaches found in previous work on the uncanny valley of mind.

1.1 Robots in the Uncanny Valley (of Mind)

The prominent *uncanny valley* hypothesis (Mori, 1970; for recent reviews see Diel & MacDorman, 2021; Mara et al., 2022) states that responses to robots get more favorable with increasing human-like appearance until a steep drop is observed for highly human-like machines—prompting eeriness, disgust, or fear. At the high end of the human likeness dimension, user responses are expected to turn positive again, reaching the most positive levels for perfectly human-like robots. This non-linear relationship between human likeness and user responses should be even stronger for moving than for static entities (Mori, 1970). Research that used morphed images found support for the uncanny valley hypothesis (e.g., Lischetzke et al., 2017; MacDorman & Ishiguro, 2006; Mathur & Reichling, 2009, 2016) but this line of research was criticized for the lack of external validity (Diel et al., 2022; Palomäki et al., 2018). A recent review and meta-analysis (Mara et al., 2022) demonstrated that higher scores on human likeness were absent in experiments that used realistic human-like robots. Moreover, scholars have raised doubt about the proposed curvilinear relationship (e.g., Hanson, 2005; MacDorman et al., 2009; Poliakoff et al., 2013), instead suggesting alternative hypotheses such as an *uncanny cliff* (Bartneck et al., 2007a). In sum, current evidence in traditional uncanny valley research suggests that increasing human likeness up to a certain point leads to negative user evaluations, while it is still uncertain whether and how these will improve once very human-like robots are available to be examined.

Beyond research on robots' visual appearance, newer research is focused on the influence of mind attributed to a robot on user responses (Appel et al., 2020; Dang & Liu, 2021; Gray & Wegner, 2012; Müller et al., 2020; Stein & Ohler, 2017). According to Gray et al. (2007), mind can be distinguished into *experience* (i.e., the ability to feel emotions, have a personality and consciousness) and *agency* (i.e., self-control, morality, memory, recognition, planning, communication, and thinking). The authors further reported that people use both dimensions to characterize the mind of healthy human adults—and might also be comfortable with assigning them to certain animals or mythological entities. While a basic degree of anthropomorphism and mind attributed to robots was found to have a positive influence on trust (Waytz et al., 2014), morality (Young & Monroe, 2019), or usefulness (Liu & Liao, 2021), ascribing human-like mind in terms of agency and experience (e.g., based on a complex artificial intelligence) to computers, smart speakers, or robots resulted in notable discomfort and apprehension (e.g., Appel et al., 2020; Brink et al., 2019; Gray & Wegner, 2012; Kang & Sundar, 2019; Taylor et al., 2020; Zafari & Koeszegi, 2020). For example, Appel et al. (2020) used text vignettes in which they described a new generation of robots. In a series of experiments, they showed that robots equipped with mental capabilities evoked higher eeriness than simple-tool robots. Not only did a robot with experience evoke the highest aversion, but also a robot with agency was rated less positive than a simple tool robot. This pattern of results was unaffected by the ascribed gender of the robot and attenuated (but not nullified) by introducing the robot to serve in a nursing environment.

Even though the relative contributions of the two mind dimensions (experience and agency) to these outcomes are a matter of on-going academic debate (e.g., Otterbacher & Talias, 2017; Yam et al., 2020), scholars warn that the growing mental prowess of machines can be detrimental to their success—pushing them into an *uncanny valley of mind* (Stein & Ohler, 2017). According to recent evidence (Appel et al., 2020; Gray & Wegner, 2012; Kang & Sundar, 2019) and in line with the evidence from the traditional uncanny valley research, it remains unclear how these negative responses to robots with mind could be overcome.

1.2 Interplay of Mind and Empathy

The current project scrutinizes a possible boundary condition and solution to users' aversion towards robots with mind: Evoking user empathy as a protective mechanism against negative user evaluations (for introductions to empathy in human-robot interaction, see Malinowska, 2021; Vanman & Kappas, 2019). Numerous efforts in interpersonal research (Batson et al., 1997; Cao, 2013; Kaseweter et al., 2012; Lotz-Schmitt et al., 2017; Meuwese et al., 2017) provide evidence that there is a positive association between empathy and positive reactions and positive attitudes towards the target of one's empathic response. Meta-analytical evidence (McAuliffe et al., 2020) suggests that representing an entity in a harmful situation is a suitable way to induce human empathy. Relocating this knowledge to human-robot interaction, when thinking about empathy for robots, the *computers as social actors* paradigm (Reeves & Nass, 1996) may readily come to mind. According to Reeves and Nass (1996), people tend to apply social norms from human-human interactions to their encounters with digital technology, not least including robots (Bartneck et al., 2007b; Eyssel & Hegel, 2012; Kahn et al., 2012; Lee et al., 2006). Thus, even if observers of robots are aware that these machines are inanimate objects, they tend to fall back on the same interaction scripts that they have acquired in real social interactions.

Building upon this framework, studies have uncovered that artificial entities are often ascribed human-like gender attributes (Nass et al., 1997), skills (Nass & Moon, 2002), and personality characteristics (Moon & Nass, 1996; Nass & Lee, 2001). Moreover, it has been shown that people not only feel empathy for other humans and animal species (e.g., de Vignemont & Singer, 2006; Young et al., 2018) but may also empathize with robotic machines (Horstmann et al., 2018; Mattiassi et al., 2021; Rosenthal-von der Pütten et al., 2013), wherein empathy is particularly likely to occur when similarities between humans and robots are salient (de Vignemont & Singer, 2006; Riek et al., 2009). In line with this evidence, people seem to be particularly prone to empathizing with machines that demonstrate some level of experience (Choi et al., 2021; Nijssen et al., 2019). As such, we underscore the distinct role of advanced and human-like robot minds for empathic user responses. Considering that empathy has further been described as a promising way to prevent technology aversion (Diel & MacDorman, 2021; Gonsior et al., 2011) eliciting empathy might indeed emerge as a key protective factor against the uncanny valley of mind.

At this point, it should be mentioned that the term empathy used in our study refers to cognitive empathy, describing a human's ability to understand another entity's situation and feelings and being able to take its perspective (Cuff et al., 2016)—while still being aware of the self-other distinction (de Vignemont & Singer, 2006). Following the interpersonal literature, the computers as social actors perspective, and evidence from human-robot interaction research, we propose that depicting robots in physically harmful situations should be a particularly effective way of triggering user empathy—even more so if the robot is perceived to have mind.

1.3 The Current Project

Taken together, we pursue a novel approach to counteract the uncanny valley of mind by proposing user empathy as a particularly powerful psychological state that can make robots with complex minds seem more likeable. By these means, we expand upon earlier research on the interplay of mind and empathy, which has typically explored perceived robot mind as a dependent variable (Küster & Swiderska, 2021). In contrast to this prior approach, robot mind is used as one of our independent variables, accompanied by robot harm: A commonly applied method to induce empathy with machines is to depict them as they fail or are intentionally harmed by humans (Bršćić et al., 2015; Menne & Schwab, 2018; Rosenthal-von der Pütten et al., 2013). Such depictions were further connected to more positive user evaluations than presenting robots in neutral situations (Gompei & Unemuro, 2015; Mirnig et al., 2017; Ragni et al., 2016). This leads us to propose a mediation model, assuming that a harmful situation evokes higher empathy, which in turn leads to higher likeability.

H1a: A robot shown in a harmful situation evokes more likeability than a robot shown in a neutral situation (main effect of the situation).

H1b: This effect is mediated by participants' empathy.

Next, we consider potential interaction effects with the robot's mind, serving as a moderator variable. Matching the well-established fact that empathy is fostered by perceived self–other similarity (Cikara et al., 2011; Hasson et al., 2018), studies from the field of social robotics revealed that attributions of mind to machines prompt stronger empathic reactions (Choi et al., 2021; Lucas et al., 2016; Nijssen et al., 2019; Yam et al., 2020). As such, we hypothesize that robots with complex mental abilities should evoke stronger empathy in harmful situations than robots that appear as simple working tools without mind. Correspondingly, this experience should translate to improved likeability and, thus, to a reduction or potentially even nullification of the uncanny valley of mind effect. Accordingly, we propose an interaction effect between the independent variable *situation* and the moderator variable robot mind on both the dependent variable likeability and the mediator variable empathy:

H2a: If shown in a neutral situation, a robot without mind evokes more likeability than a robot with mind. This effect is reduced, nullified, or reversed if the robot is shown in a harmful situation (interaction effect).

H2b: This effect is mediated by participants' empathy (moderated mediation).

2. Experiment 1

2.1 Method

An online experiment was conducted to test our preregistered hypotheses (https://aspredicted.org/td2c7.pdf). We randomly assigned participants to conditions that either introduced a robot with or without mind (Factor 1: robot mind), before depicting the machine in a neutral or harmful situation (Factor 2: situation). Thus, the study followed a 2×2 between-subjects design.

2.1.1 Participants

In prior research, the effect of mind (experience vs. tool condition) on eeriness amounted to $d = 1.05$ (Appel et al., 2020, Experiment 2). The lower bound of the 60% confidence interval (Perugini et al., 2014) was $d = 0.89$. We expected that the effect of the robot introduction could be smaller in our design, given that we presented our robot videos after the introduction and before the dependent variables. In consequence, we determined the focal effect size to be $d = 0.60$. A power analysis with G*Power (Faul et al., 2007) left us with an aspired sample size of 64 for a two-group main effect (two-tailed independent t-test, power = .80, alpha-error-probability = .05). To account for the more complex design and the power needed to identify an interaction effect, we multiplied this sample size by the factor eight (Giner-Sorolla, 2018; Simonsohn, 2014), leading to a proposed sample size of 512. We invited 650 U.S.-American residents from the *MTurk* online participant pool (selection criteria: hit approval rate > 97%, hits > 1000) to have a buffer if careless responding occurred.

Of the 650 completions, 20 participants did not have sufficient English skills, as indicated by a control question, and were therefore not included in our statistical analyses (Kennedy et al., 2020). While a self-report attention check was answered positively by all remaining participants, six individuals showed large (> ±3 years) deviations when asked twice about their age, leading to their removal from the data. Moreover, seven participants were excluded because their participation time was lower than 150 seconds. As treatment checks, participants were asked to indicate whether they had been introduced to a robot that is a simple tool or a robot that is characterized by mind and personality—and whether the robot in the video had been harmed or not harmed by a human. Based on that data, an additional 49 participants did not describe the correct type of robot (43) or the depicted situation (6) and were therefore excluded, as well as nine participants who did not recognize an item shown in our stimuli as an additional attention check. The final sample consisted of 559 participants (270 female, 287 male, 2 non-binary or no answer) with an average age of 41.43 years (*SD* = 12.04, ranging from 22 to 78 years). In Experiment 1, gender was equally distributed across conditions[1], $\chi^2(3, N = 557) = 6.11$, $p = .106$, $\varphi = .11$, as was the case for age $F(3, 555) = 1.46$, $p = .224$. Most of the sample described themselves as White American (77.64%), followed by Black/African American (9.12%), Asian American (5.90%), and Hispanic/Latino (5.01%).

2.1.2 Stimuli and Procedure

After giving informed consent, participants were exposed to the experimental manipulations. As is common in the research field, our first experiment made use of vignette texts to introduce participants to a robot with different degrees of mental sophistication. One text described the robot as a *tool robot* managing its everyday tasks without any complex mental capabilities. In contrast, the other text described the robot as being able to feel and think based on complex artificial intelligence technology (see online supplement for the full materials). The design of this manipulation followed prior research (Appel et al., 2020; Gray & Wegner, 2012; Ward et al., 2013), as we hoped to uncover consistent evidence of the uncanny valley of mind. In a significant deviation from previous work, however, we next presented participants with brief videos (50 seconds) showing a humanoid robot (Atlas model by Boston Dynamics), claiming it to be the machine from the vignette texts. In the first condition of this manipulation (*harmful situation*), we showed the robot as it attempted to pick up a box but was repeatedly stopped by an adult man, who used a hockey stick to push

[1] The two participants who answered “diverse” or “no answer” when being asked for their gender were not included in this analysis.

the box out of the robot's reach. At the end of the video, the robot was further pushed to the ground by the human with the hockey stick that had pushed the boxes away in the first half of the video. The last scene shows the robot lying on the floor after it has been knocked down by the human with that stick. The second video (*neutral situation*) showed the robot successfully putting boxes into a shelf and walking around on several surfaces. After watching the assigned video, participants rated the likeability of the shown robot as well as their experienced empathy. Additionally, they answered several attention check and control items before providing sociodemographic data. In the end, participants were thanked and debriefed. The MTurk participants were compensated with 1 USD for their participation in our experiment, which took about five minutes. The internal review board at at the Human-Computer-Media Institute of the Julius-Maximilians-University of Würzburg approved the experiment (reference 010721). On the final pages, we asked participants whether or not they had seen parts of the video before. A large majority of participants (93.12%) reported not having watched parts of the video prior to this study.

2.1.3 Measures

Likeability. To assess the robot's likeability, we used the five likeability items of the Godspeed Questionnaire (Bartneck et al., 2009). The semantic differential scales ranged from 1 to 5, M = 3.77 (SD = 0.82), Cronbach's α = .93.

Empathy. We assessed the participants' state empathy towards the robot with an ad-hoc scale based on the work by Oswald (1996). The three items "empathic", "softhearted", and "compassionate" were presented on a 5-point scale ranging from 1 *(does not at all describe how I feel)* to 5 (*describes how I feel extremely well*), M = 2.90 (SD = 1.44), Cronbach's α = .98.

2.2 Results

We first tested the hypothesized main effect (H1a) and the interaction effect (H2a) pertaining to robot likeability using an analysis of variance (ANOVA). A small but significant main effect of the situation was found, $F(1, 555) = 4.27$, $p = .039$, $\eta_p^2 = .01$, supporting H1a. Likeability was higher for the robot in the harmful situation ($M = 3.84$, $SD = 0.86$) than for the robot in the neutral situation ($M = 3.70$, $SD = 0.76$). Neither a main effect of robot mind, $F(1, 555) = 0.49$, $p = .483$, $\eta_p^2 < .01$, nor an interaction effect between both factors was observed, $F(1, 555) = 0.22$, $p = .639$ $\eta_p^2 < .01$. Thus, H2a has to be rejected based on our data. The main descriptive results of both experiments are displayed in Table 1. Additionally, the results for Experiment 1 are illustrated in Figure 1.

Table 1

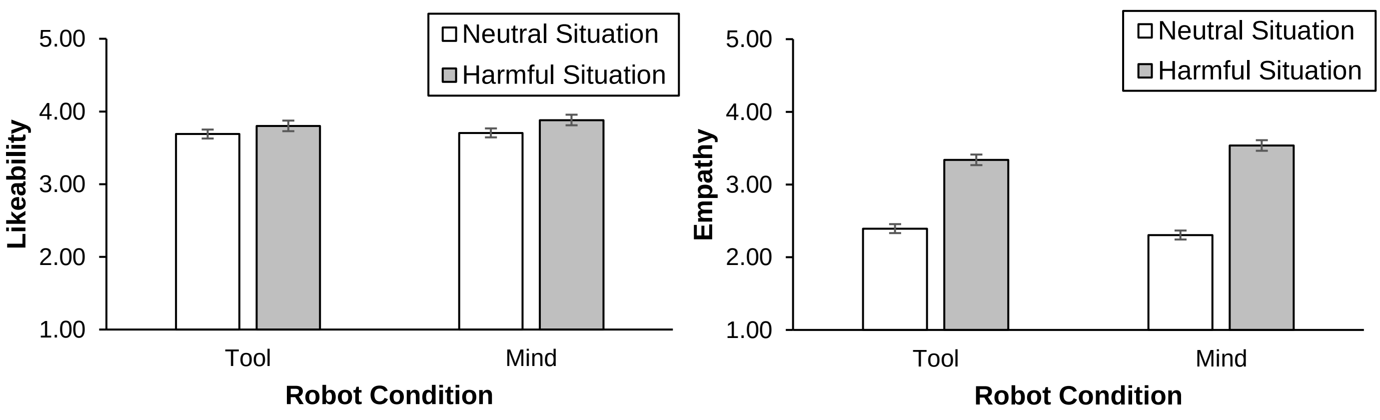
Likeability and Empathy Means and Standard Deviations for Both Experiments

	Likeability				Empathy			
	Mind		Tool		Mind		Tool	
Situation	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Experiment 1								
Harmful	3.88	0.88	3.80	0.84	3.54	1.39	3.34	1.44
Neutral	3.71	0.81	3.69	0.71	2.31	1.22	2.39	1.27
Experiment 2								
Harmful	3.51	0.89	3.33	0.92	2.83	1.28	2.25	1.26
Neutral	3.71	0.86	3.43	0.70	2.57	1.18	1.91	0.96

Note. Sample sizes Experiment 1: Mind–Harmful: *n* = 153, Mind–Neutral: *n* = 146, Tool–Harmful: *n* =128, Tool–Neutral: *n* = 132. Sample sizes Experiment 2: Mind–Harmful: *n* = 98, Mind–Neutral: *n* = 100, Tool–Harmful: *n* = 93, Tool–Neutral: *n* = 105.

Figure 1

Likeability and Empathy Means (with Standard Errors of the Mean) Depending on Robot Condition and Situation in Experiment 1



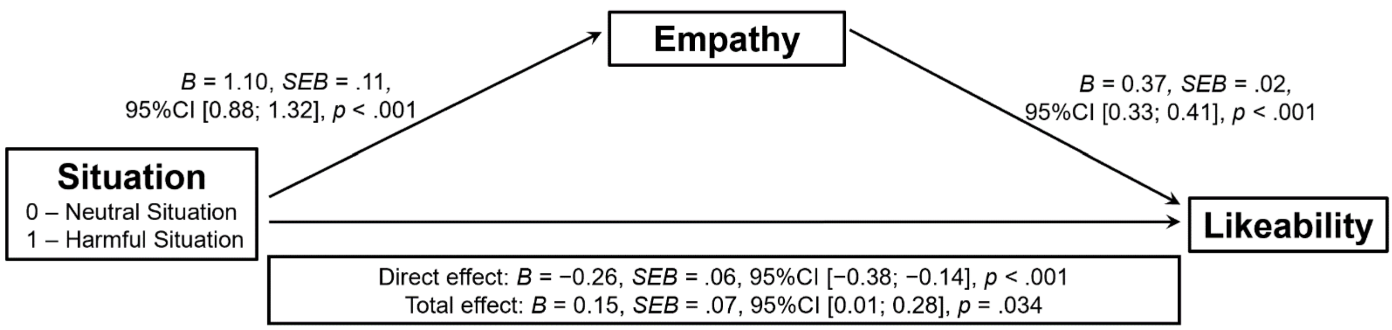
Note. Error bars represent ± 1SE.

Focusing on the variable empathy, a second ANOVA yielded a significant main effect of the situation, $F(1, 555) = 92.65$, $p < .001$, $\eta_p^2 = .14$. Empathy was higher for the robot in the harmful situation ($M = 3.45$, $SD = 1.41$) than for the robot in the neutral situation ($M = 2.35$, $SD = 1.25$). In contrast to this, neither a main effect of robot mind, $F(1, 555) = 0.23$, $p = .628$, $\eta_p^2 < .01$, nor an interaction effect could be uncovered, $F(1, 555) = 1.60$, $p = .207$, $\eta_p^2 < .01$.

Proceeding to the relationship between the mediator and dependent variables, we found a significant correlation between empathy and likeability, $r(557) = .59$, $p < .001$. Furthermore, the mediation model formulated in H1b was supported by our data. Using the PROCESS macro for SPSS software (Hayes, 2018), a significant indirect effect was observed, $B = 0.40$, bootstrapped $SE = 0.05$, bootstrapped 95% CI [0.31, 0.51]. Figure 2 presents an overview of the parameters uncovered in the mediation analysis. In addition to the indirect effect, we found a negative direct (i.e., residual) effect of the situation on likeability.

Figure 2

Results of the Mediation Model of Experiment 1



In line with the reported non-significant interaction effects of the ANOVAs (see Figure 1), no significant index of moderated mediation (= 0.11, bootstrapped *SE* = 0.08) was observed, bootstrapped 95% CI [−0.06, 0.27]. Thus, H2b was not confirmed.

2.3 Discussion

Consistent with earlier research (Menne & Schwab, 2018; Rosenthal-von der Pütten et al., 2013; Seo et al., 2015), our results reveal that humans indeed feel empathy for robots that are presented in a physically harmful situation. Furthermore, the increase in empathy for the robot significantly predicted higher likeability, leading to statistically relevant indirect and total effects. This suggests that making observers empathize with a human-like machine may indeed hold strong merit for efforts to increase technology acceptance.

To our surprise, however, the robot with mind was rated as likeable as the robot without mind in the neutral situation in our experiment—indicating that participants were not particularly dismissive of a robot with advanced abilities. Furthermore, a negative direct (i.e., residual) effect of the depicted situation on likeability occurred, in that the harmful situation made the robot seem less likeable once empathy was taken out of the equation. Apparently, this implies that some unobserved factors led participants to feel less positive about the harmed robot—although this effect was ultimately overridden by the positive indirect effect via increased empathy.

In summary, our results indicate that the situation in which the robot was presented affected our participants' evaluations much more than its alleged mental capacities. This suggests that the interplay between robot mind and empathy might be less pronounced than previously assumed. At the same time, we cannot rule out that our results depended on the way our stimuli were perceived by participants. Since psychological insight shows that visual cues are usually processed with higher priority (Hernández-Méndez & Muñoz-Leiva, 2015;

Koć-Januchta et al., 2017; Navon, 1977), it is possible that the video manipulation (of the situation) affected our participants more than the text manipulation (of robot mind). Due to this methodological limitation, we decided to conduct a second experiment based on video materials that manipulated mind and situation at the same time.

3. Experiment 2

Surprisingly, one of our main assumptions—a robot with human-like mind should be ascribed less likeability than a robot without mind in a neutral situation—was not supported in the first experiment. In order to rule out the possibility that this was due to the modality of the chosen manipulation, we conducted a second experiment with a focus on video recordings for the realization of our conditions. Specifically, we created four videos that manipulated both the situation (harmful vs. neutral) and the robots' mind (robot with mind vs. robot without mind). Again, the study followed a 2×2 between-subjects design and was preregistered in terms of hypotheses, materials, and planned analyses (https://aspredicted.org/cs63u.pdf). Regarding our specific assumptions, we pursued the same propositions as in Experiment 1.

3.1 Method

3.1.1 Participants

To calculate the required sample size, we once more relied on the effect size reported by Appel et al. (2020, $d = 1.05$). The lower bound of the 60% confidence interval (Perugini et al., 2014) was $d = 0.89$ and used for the power analysis as written and visual stimuli were now combined into a single treatment. Using G*Power software (Faul et al., 2007), we obtained an aspired sample size of 42 for a two-group main effect (two-tailed independent t-test, power = .80, alpha probability = .05). To account for the more complex design and the power needed to identify an interaction effect, we multiplied this sample size with the factor 8 (Giner-Sorolla, 2018; Simonsohn, 2014), leading to a proposed sample size of 336. Yet, to guarantee enough power in the case of some careless responding, we asked 400 persons of the *Prolific* participant pool to participate in the online experiment.

Of the 400 completions, a single participant did not have sufficient English skills and was therefore removed from our statistical analyses (Kennedy et al., 2020). Another three participants had large (> ±3 years) deviations when asked twice about their age. A control question on the general topic of the study was answered correctly by all participants and no individuals had to be excluded based on their answering duration. Lastly, we used two treatment check items asking participants whether the robot had been damaged or not and whether it was described with or without elaborate mental abilities. However, since we retrospectively noticed some problems with the wording of these items, we decided against using them as an exclusion criterion (as initially planned)[2]. To compensate for this, an additional test of our materials' validity was carried out, yielding positive results (please see section 3.1.3).

The final sample consisted of 396 participants (232 female, 154 male, 10 non-binary or no answer) with an average age of 38.66 years (SD = 14.06, ranging from 18 to 82 years). As in Experiment 1, gender was equally distributed across conditions[3], $\chi^2(3, N = 386) = 5.62$, $p = .131$, $\varphi = .13$, as was the case for age $F(3, 329) = 0.43$, $p = .733$. Most participants described themselves as White American (80.30%), followed by Asian American (6.82%), Black/African American (5.56%), and Hispanic/Latino (4.29 %).

3.1.2 Stimuli and Procedure

Four different videos were created to manipulate the robot's mind and exposure to harm. Each clip had a duration of approximately 80 seconds, including five scenes each (see online supplement for the full screenplays). Again, the videos showed the humanoid robot

---

[2] See the online supplement for further details about the in- and exclusion of participants.

[3] Ten participants who answered "diverse" or "no answer" when being asked for their gender were not included in this analysis.

Atlas in a laboratory setting, albeit covering a much broader range of harmful (harassment by a human confederate with a stick, robot knocked down with a stick) or harmless situations (simple working tasks, human pushing a box with a stick). Moreover, addressing our second experimental factor robot mind, a female narrator described the robot and its capabilities. Subtitles were added to make sure that this information remained salient even if participants felt inclined to focus more on the visuals. The robot was described either as a sophisticated co-worker that could act independently of human commands and feel some forms of emotions due to its neural network technology—or as a simple tool that had to be explicitly programmed for all relevant tasks.

Following the presentation of the randomly assigned video in Experiment 2, participants rated likeability and empathy, answered several attention check items, and questions about their sociodemographic background. Again, most of the sample (91.14%) had not watched the original parts of the video prior to the study. We thanked them for their participation and debriefed our participants. The Prolific participants were compensated with 1 USD for their participation in our experiment, which took about five minutes. The internal review board at the Human-Computer-Media Institute of the Julius-Maximilians-University of Würzburg approved the experiment (reference 091221).

3.1.3 Additional Data on the Success of the Experimental Manipulation

To test the successful creation of the videos inducing the manipulation of robot mind and the harmful situation, an independent online sample of 423 participants was exposed to the stimuli (see online supplement S6). Like the sample in the main study, the additional sample was recruited via Prolific (M_{age} = 43.55, SD_{age} = 14.81, 225 male). We relied on four items by Gray and Wegner (2012) to test perceived robot mind (M = 3.14, SD = 1.76, Cronbach’s α = .87) and on four self-created items (M = 2.39, SD = 1.83, Cronbach’s α = .95) inspired by Rosenthal-von der Pütten et al. (2013) to test the perception of harm in the shown situation (see online supplement for details). These items were presented on 7-point scales ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).

Results revealed that the robot described to have human-like mental capabilities was perceived to have mind to a much larger extent (M = 4.31, SD = 1.50) than the tool robot (M = 2.03, SD = 1.71), $t(421) = 17.48$, $p < .001$, d = 1.70. The shown situation also yielded the expected effect: The robot in the harmful situation was perceived to be harmed more (M = 3.60, SD = 1.94) than the robot in the neutral situation (M = 1.32, SD = 0.73), $t(421) = 16.32$, $p < .001$, d = 1.59. This additional data based on an unrelated sample corroborated our assumption that the videos lead to the intended effects in terms of participants' perceptions of harm and mind.

3.1.4 Measures

Likeability. To assess the robot's likeability, we again used the five-point likeability scale of the Godspeed Questionnaire (Bartneck et al., 2009), M = 3.50 (SD = 0.85), Cronbach's α = .92.

Empathy. We used the same items by Oswald (1996) as described in the first experiment, M = 2.39 (SD = 1.22), Cronbach's α = .97.

3.2 Results

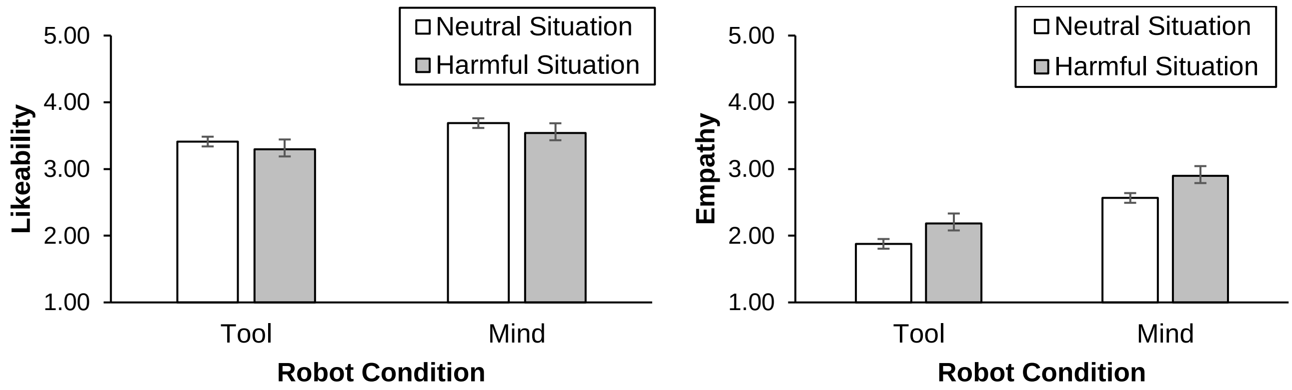
Since we retained the hypotheses from the first experiment, all analyses follow the same data analysis plan (see Figure 3 for the main results). Considering the dependent variable likeability, no main effect of the situation was found, $F(1, 392) = 2.97$, $p = .086$, $\eta_p^2 = .01$. In contrast to H1a and the results of Experiment 1, likeability scores for the robot in the neutral situation (M = 3.56, SD = 0.79) and the robot in the harmful situation (M = 3.42, SD = 0.91) did not differ. However, a significant main effect of robot mind could be observed, $F(1, 392) = 7.60$, $p = .006$, $\eta_p^2 = .02$. The robot with mind was rated as more likeable (M = 3.61, SD = 0.88) than the robot without mind (M = 3.38, SD = 0.81). The interaction effect did not reach statistical significance, $F(1, 392) = 0.30$, $p = .587$, $\eta_p^2 < .01$. The direction of the mean difference and the absence of an interaction effect were in contrast to our expectations. Thus, H2a was rejected.

Proceeding to the investigation of the outcome empathy, a main effect of the situation was found, $F(1, 392) = 6.38$, $p = .012$, $\eta_p^2 = .02$. Empathy was higher for the robot in the harmful situation ($M = 2.55$, $SD = 1.30$) than for the robot in the neutral situation ($M = 2.23$, $SD = 1.12$). Additionally, the ANOVA revealed a main effect of robot mind, $F(1, 392) = 27.65$, $p < .001$, $\eta_p^2 = .07$. The robot with mind ($M = 2.70$, $SD = 1.24$) evoked higher empathy than the robot without mind ($M = 2.07$, $SD = 1.12$). Yet, the interaction term did not reach statistical significance, $F(1, 392) = 0.01$, $p = .768$, $\eta_p^2 < .01$.

Finally, we focused on the interplay of our outcome measures as well as a potential mediation (Figure 4). Again, empathy and likeability were found to be significantly correlated, $r(394) = .57$, $p < .001$. Moreover, the mediation model formulated in H1b was supported by our data. As hypothesized, a significant indirect effect occurred, $B = 0.12$, bootstrapped $SE = 0.05$, bootstrapped 95% CI [0.03, 0.23]. In line with the reported non-significant interaction effects of the ANOVAs, no significant index of moderated mediation (= 0.03, bootstrapped $SE = 0.10$) was observed using the SPSS macro PROCESS (Hayes, 2018), bootstrapped 95% CI [−0.17, 0.23]. Thus, H2b was again not confirmed by our data. As in Experiment 1, we observed a negative direct effect of the situation on likeability, indicating that a robot presented in a harmful situation evoked less likeability than a robot presented in a neutral situation if the effect on empathy is statistically controlled (see Figure 4).

Figure 3

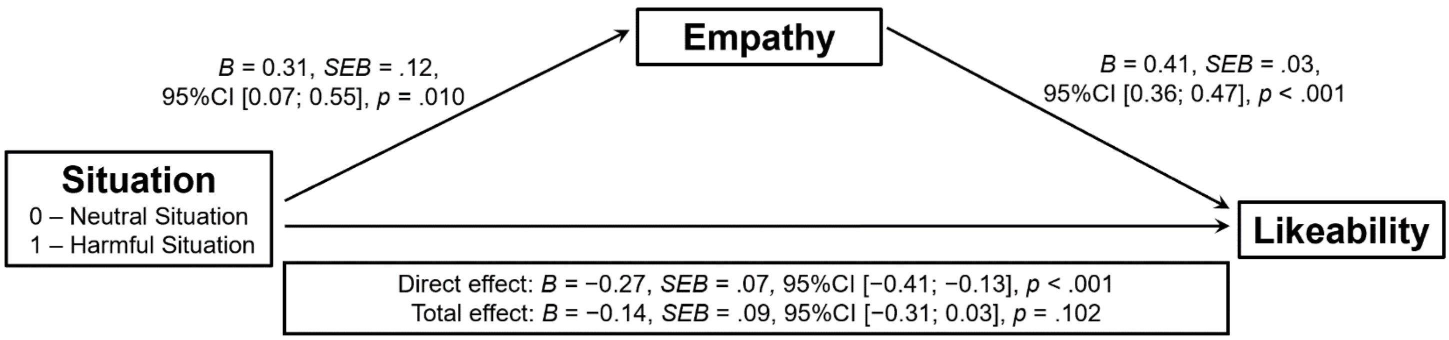
Likeability and Empathy Means (with Standard Errors of the Mean) in Dependence of Robot Condition and Situation in Experiment 2



Note. Error bars represent ± 1SE.

Figure 4

Results of the Mediation Model of Experiment 2



3.3 Discussion

By presenting one out of four videos that combined the manipulation of robot mind and harm into a single treatment, we made sure that neither manipulation was able to override the other from an attentional point of view. Furthermore, we employed a broader range of situations, extending the videos of Experiment 1 in duration and complexity, and offered a more detailed description of the robot’s supposed mind to strengthen the rigor of our manipulations. Consistent with Experiment 1, our empirical efforts showed that observing a human-like robot in a harmful situation prompted empathy, which led to more liking as part of a significant indirect effect. Contrary to Experiment 1, however, the harmful situation did not evoke higher likeability than the neutral situation in terms of a total (main) effect. This finding can be attributed to the residual negative effect of presenting the robot in a harmful situation; with positive indirect and negative direct effect again opposing each other. Since the mediation effect was smaller this time around, the total effect ultimately turned out insignificant.

Regarding the role of the robot’s mind, our second experiment yielded a slightly different picture compared to the first experiment. Now, a robot with complex mental abilities was not only liked as much as its simpler counterpart in neutral situations but actually preferred by our participants. Similar to the surprising negative residual effect of the shown situation on likeability, we believe that these unexpected findings warrant deeper discussion.

4. General Discussion

For several centuries, humans have indulged in the idea of co-existing with advanced robotic machinery, as countless works from literature and the arts vividly illustrate. Be it the deceptively human-like Olympia in E.T.A. Hoffmann’s novel “The Sandman” (1816/2012) or the sophisticated androids in modern science fiction movies: In most artistic visions, robots are conceived as highly anthropomorphic beings, which are equipped with impressive mental competence. As real-world technology has started to catch up with fiction, however, researchers eventually noted that people tend to become apprehensive once a machine’s mind resembles the ‘human’ way of thinking or feeling too closely (e.g., Gray & Wegner, 2012). This presents a notable problem, as a whole industry sector currently focuses on the creation of increasingly intelligent technologies.

Due to the fact that people are able to empathize with human-like machines (e.g., Menne & Schwab, 2018; Rosenthal-von der Pütten et al., 2013), we explored the idea that empathy for robots might help to mitigate the uncanny valley of mind. In line with our hypotheses, both experiments revealed that presenting a robot in a physically harmful situation leads to higher empathy than a depiction without harm. In turn, this response predicted higher likeability, culminating in a significant mediation effect—consistently in both studies. As such, we want to underscore empathy as a highly relevant mechanism to improve users' evaluation of human-like technology. Also, since our second experiment indicated that robots with mind could indeed trigger stronger empathic reactions than simple tool robots, social cognitive processes may hold particular relevance to ensure the approval of such innovations.

At the same time, it is important to point out that in both experiments, the sizes of the respective effects turned out rather moderate or small, so that they should be interpreted with the appropriate caution. Further, a rather surprising takeaway from our empirical efforts emerged as we did not observe stronger aversion towards the mind robot in neutral situations—in other words, no uncanny valley of mind effect. Especially Experiment 2 showed that the robot that was described as having more complex abilities was perceived more positively than its tool-like counterpart, thus contradicting previous studies.

4.1 Uncanny Valley of Mind Revisited

The uncanny valley of mind has primarily been approached with text vignettes about innovative machinery that may exist in the future (e.g., Grundke et al., 2022; Taylor et al., 2020). With our combination of texts and videos (Experiment 1) or videos (Experiment 2) as stimuli, we could not replicate the negative effect of perceiving minds in machines. Several possible reasons need to be noted. First of all, the robot in our video clips did not actually demonstrate any of the advanced capabilities that were described in our complementary vignettes or narrations; instead of "thinking" or "feeling," it was mostly shown moving around or executing physical tasks. In all probability, our decision to only make use of preexisting, natural video materials may have limited our ability to show particularly aversive or eerie situations in this regard. As previous literature has highlighted that people might be especially wary of new technology acquiring *social* and *emotional* abilities (e.g., Appel et al., 2020; Stein & Ohler, 2017), it seems necessary to follow up on the presented work with materials that actually depict human-robot interactions as well as affective reactions by robots. For the current project, however, we cannot rule out that the robots in the presented video clips ultimately appeared much simpler than the hypothetical machines described in other studies (Appel et al., 2020; Gray & Wegner, 2012)—thus falling short of the uncanny valley of mind.

Another important clue to the obtained lack of findings might be the undeniable importance of a robot's appearance for people's evaluation (i.e., the classical uncanny valley). Explicitly comparing the impact of visual and mental aspects, recent literature suggests that participants tend to focus more on a machine's design than its abilities when evaluating its eeriness (e.g., Ferrari et al., 2016; Stein et al., 2020). Arguably, this matches the broader psychological understanding of human information processing, as visual cues often take immediate priority over other available information (e.g., Hernández-Méndez & Muñoz-Leiva, 2015; Koć-Januchta et al., 2017; Navon, 1977). Further complicating matters, interaction effects may ensue, as different robot appearances (e.g., having a face or no face) might modulate the corresponding attributions of mind. Atlas, the robot shown in our video materials, only possesses rudimentary human-like cues—i.e., an upright, bipedal stature but no facial features and falls in a cluster of robots evaluated to be very mechanical, neither indicating high likeability nor high threat when presented without context (Rosenthal-von der Pütten & Krämer, 2014). Due to the robot's mechanical appearance, it might have been more difficult for participants to imagine its complex mental abilities; a notable difference from previous vignette studies that often left the specific look of the machine entirely to participants' imagination.

In summary, we do not consider our results pattern as a rebuttal to uncanny valley of mind theory. Instead, we present our findings as proof that further research with different types of robots, situations, and modalities is required to examine the boundaries of this phenomenon. If possible, this should involve not only natural video materials but also actual human-robot interactions, both in laboratories and in the field (see Mara et al., 2021, for a comparison between presentation modes).

4.2 Unaccounted Effects of Robot Competence

Another unexpected finding in our project was a negative residual effect of presenting the robot in a harmful situation on participants' likeability ratings: Across both experiments, individuals actually liked robots in harmful situations *less* than robots in neutral situations once we controlled for the indirect influence of empathy. This is somewhat surprising, keeping in mind previous evidence that reported positive user impressions after robot failures (Mirnig et al., 2017; Ragni et al., 2016; Salem et al., 2013). Nevertheless, potential explanations for our observations are offered by both literature and practical considerations. A robot that is unable to withstand physical harm can be easily interrupted during the fulfillment of its tasks, which hinders the robot from fulfilling its task in a competent manner and creates notable problems for the work context. In line with this thought, studies show that people are rather intolerant of algorithm failure (Dietvorst et al., 2015), expect competent service from technology (Waytz et al., 2014), and perform worse in response to a failing machine (Robinette et al., 2017; Salem et al., 2013; van den Brule et al., 2014). Similarly, the recent work by Chen et al. (2021) shows that customers are less forgiving if errors were made by an incompetent self-service technology instead of human employees. This emphasizes the high expectations people have towards modern-day machinery, whose raison d'être is to perform tasks reliably, to be competent, and to assist humans (Broman & Finckenberg-Broman, 2017; Brooks et al., 2016; Horstmann & Krämer, 2019).

In addition to that, a robot that fails to perform its tasks is not only potentially useless to co-workers but could also put people in real danger. In our videos of harmful situations, one push suffices to topple over the robot Atlas (height: 1.50 m, weight: 196 lbs = 89 kg) and make it fall to the ground. Considering the machine’s dimensions, it could have easily hit and hurt another person or at least another object by falling. We assume that if people have to face the possibility of being physically injured by a robot at any time, they might come to evaluate it less favorably—even if they may simultaneously empathize with it.

In a similar vein, we note that robot incompetence (i.e., a machine that cannot cope with its tasks or unforeseen circumstances) will cause additional work stress to human users (e.g., Fallatah et al., 2019; Michalos et al., 2015), which might have further informed our results. For example, humans would have to take care of error messages, make adjustments to the robot’s code to ensure its continued operation, or pick it up after a fall.

Lastly, we argue that the negative residual effects of watching a harmed machine could actually be interpreted as an indication of the uncanny valley of mind after all—considering that “to err is human”—and additionally, lacking competence in several situations is also human. While qualitative data on participants’ experiences would be needed to consolidate this idea, we have come under the impression that the harmful situation made the robot appear much more human-like than the neutral counterpart, in which the machine might have appeared artificial or superior. Thus, we consider it likely that the situations we showed eventually contributed to our mind manipulation as well, providing participants with a way to perceive more or less (mental) human likeness in the robot Atlas.

4.3 Limitations and Future Research

In addition to the limitations already stated above, several other aspects limit the generalizability of our findings. In both experiments, a non-negligible amount of people had problems correctly recognizing their experimental condition and had to be excluded, especially in the condition “tool robot, harmful situation” in Experiment 2. While reconducting the analysis for both experiments *with* or *without* the concerning participants yielded the same results pattern (see online supplement), this implies that future studies might benefit from even stronger, more explicit video materials and clearer treatment check items. Based on the results of an independent sample, we nevertheless assume that our general manipulation of situation and mind was successful. As stated above, the depiction of robots with more human-like features (e.g., human faces) in more social or emotional situations should also provide a meaningful next step for this line of research.

Of course, ethical concerns are an important issue in studies showing harmful situations—but within these boundaries, meaningful modifications of our methods are welcome. While we showed physical harassment by a human user to induce empathy for the robot, future work might, for instance, portray robots harming each other (to take ingroup effects into account, see Fraune et al., 2017; Steain et al., 2019), or revolve around verbal insults or ostracism. Then again, we note that empathy may also be evoked by optimistic scenarios, so research should not only focus on negative treatments, which would also have the advantage that a robot would not first have to be harmed or damaged in order for empathy to arise. This would be desirable for financial reasons as well as for ethical reasons. Instead of harming, we encourage future research to explore empathy-inducing scenarios that can be better implemented in practice and suggest here, for example, to focus on commonalities between humans and machines, as commonalities can be a way to increase empathy for counterparts (Grover & Brockner, 1989; Heinke & Louis, 2009; Osborne-Crowley et al., 2019).

Shifting attention to the participants' side, it must be considered that the intensity of empathy also depends on each person's individual predispositions. Some people feel stronger empathy in one scenario, while others feel stronger empathy in another. As such, the individual tendency to empathize could also be an interesting variable to assess (Darling et al., 2015; Mehrabian et al., 1988). Similarly, we consider prior knowledge and technical expertise with robots to be variables of great interest. As some people may be more used to interacting with robots, or find it easier to anthropomorphize them, they might also feel a closer connection, which might translate into stronger empathy (for early and late empathy responses see, e.g., Chang et al., 2021). Notwithstanding, since the participants were randomly assigned to one of the four conditions in both experiments, we assume that the influence of individual differences between groups was controlled for (Edgington, 1996; Ferron et al., 2014; Kratochwill & Levin, 2010). As an example, we highlight the potential influence of gender differences on empathy: Even if women were found to show higher trait empathy than men (Klein & Hodges, 2001; Macaskill et al., 2002; Rueckert & Naybar, 2008), we have no indication that this influences our results since participants were randomly assigned to conditions and therefore, gender did not differ across conditions as reported in the method sections of our experiments. Moreover, in both experiments, there were at least 93 participants assigned to each condition. This is a lot more than the required number to make sure that randomization is successful so that individual differences do not differ systematically between conditions—not only in a liberal (Mittring, 2004) but also in a more conservative reading (Elliott, 2007; Lachin, 1988). Of course, diverse samples from different cultures, age ranges, and educational backgrounds will be all but needed to establish generalizability for the findings at hand.

Lastly, we would like to reiterate that conducting similar work in real settings would be most valuable. Undoubtedly, directly witnessing a robot being harmed will lead to stronger reactions than just watching a video about the procedure. As a matter of fact, previous work using live interactions showed that the emotional reactions towards robots in critical situations turned out stronger in live scenarios than for videos (e.g., Horstmann et al., 2018; Seo et al., 2015).

4.4 Conclusion

The goal of our study was to explore the role of empathy as a possibility to alleviate the uncanny valley of mind. We consistently showed that robots exposed to harm elicited stronger state empathy, which led to a higher likeability of the robot. At the same time, exposing the robot to harm elicited a negative residual influence. We assume that the residual negative influence of displaying a robot's vulnerability is due to the many complications that may arise when interacting with a vulnerable robot. We further propose that the uncanny valley of mind observed in our studies could be based on the robot's human-like imperfection, rather than descriptions of its supposed mind—an exciting perspective for future research.

References

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

Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2007a, August 26-29). Is the uncanny valley an uncanny cliff? In *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., Mubin, O., & Al Mahmud, A. (2007b, November 29-December 1). The perception of animacy and intelligence based on a robot's embodiment. In *7th IEEE-RAS International Conference on Humanoid Robots* (pp. 300-305). IEEE. https://doi.org/10.1109/ICHR.2007.4813884

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*(1), 71-81. https://doi.org/10.1007/s12369-008-0001-3

Batson, C. D., Polycarpou, M. P., Harmon-Jones, E., Imhoff, H. J., Mitchener, E. C., Bednar, L. L., Klein, T. R., & Highberger, L. (1997). Empathy and attitudes: Can feeling for a member of a stigmatized group improve feelings toward the group? *Journal of Personality and Social Psychology*, *72*(1), 105-118. https://doi.org/10.1037/0022-3514.72.1.105

Brink, K. A., Gray, K., & Wellman, H. M. (2019). Creepiness creeps in: Uncanny valley feelings are acquired in childhood. *Child Development*, *90*(4), 1202-1214. https://doi.org/10.1111/cdev.12999

Broman, M. M., & Finckenberg-Broman, P. (2017, August 10-11). Human-robotics & AI interaction: The Robotics/AI legal entity (RAiLE©). In *IEEE International Symposium on Technology and Society (ISTAS)* (pp. 1-7). IEEE. https://doi.org/10.1109/ISTAS.2017.8318980

Brondi, S., Pivetti, M., Di Battista, S., & Sarrica, M. (2021). What do we expect from robots? Social representations, attitudes and evaluations of robots in daily life. *Technology in Society*, *66*, Article 101663. https://doi.org/10.1016/j.techsoc.2021.101663

Brooks, D. J., Begum, M., & Yanco, H. A. (2016, August 26-31). Analysis of reactions towards failures and recovery strategies for autonomous robots. In *25th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 487-492). IEEE. https://doi.org/10.1109/ROMAN.2016.7745162

Bršćić, D., Kidokoro, H., Suehiro, Y., & Kanda, T. (2015, March 2-5). Escaping from children's abuse of social robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 59-66). ACM. https://doi.org/10.1145/2696454.2696468

Bryndin, E. (2020). Formation of technological cognitive reason with artificial intelligence in virtual space. *Britain International of Exact Sciences (BioEx) Journal*, *2*(2), 450-461. https://doi.org/10.33258/bioex.v2i2.222

Cameron, D., de Saille, S., Collins, E. C., Aitken, J. M., Cheung, H., Chua, A., Loh, E. L., & Law, J. (2021). The effect of social-cognitive recovery strategies on likability, capability and trust in social robots. *Computers in Human Behavior*, *114*, Article 106561. https://doi.org/10.1016/j.chb.2020.106561

Cao, X. (2013). The effects of facial close-ups and viewers' sex on empathy and intentions to help people in need. *Mass Communication and Society*, *16*(2), 161-178. https://doi.org/10.1080/15205436.2012.683928

Chang, W., Wang, H., Yan, G., Lu, Z., Liu, C., & Hua, C. (2021). EEG based functional connectivity analysis of human pain empathy towards humans and robots. *Neuropsychologia*, *151*, Article 107695. https://doi.org/10.1016/j.neuropsychologia.2020.107695

Chen, N., Mohanty, S., Jiao, J., & Fan, X. (2021). To err is human: Tolerate humans instead of machines in service failure. *Journal of Retailing and Consumer Services, 59*, Article 102363. https://doi.org/10.1016/j.jretconser.2020.102363

Choi, S., Mattila, A. S., & Bolton, L. E. (2021). To err is human (-oid): How do consumers react to robot service failure and recovery? *Journal of Service Research*, *24*(3), 357-371. https://doi.org/10.1177/1094670520978798

Cikara, M., Bruneau, E. G., & Saxe, R. R. (2011). Us and them: Intergroup failures of empathy. *Current Directions in Psychological Science*, *20*(3), 149-153. https://doi.org/10.1177/0963721411408713

Cuff, B. M., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion Review*, *8*(2), 144-153. https://doi.org/10.1177/1754073914558466

Dang, J., & Liu, L. (2021). Robots are friends as well as foes: Ambivalent attitudes toward mindful and mindless AI robots in the United States and China. *Computers in Human Behavior*, *115*, Article 106612. https://doi.org/10.1016/j.chb.2020.106612

Darling, K., Nandy, P., & Breazeal, C. (2015, August 31-September 4). Empathic concern and the effect of stories in human-robot interaction. In *24th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 770-775). IEEE. https://doi.org/10.1109/ROMAN.2015.7333675

de Vignemont, F., & Singer, T. (2006). The empathic brain: how, when and why? *Trends in Cognitive Sciences*, *10*(10), 435-441. https://doi.org/10.1016/j.tics.2006.08.008

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

Diel, A., Weigelt, S., & Macdorman, K. F. (2022). A meta-analysis of the uncanny valley's independent and dependent variables. *ACM Transactions on Human-Robot Interaction (THRI)*, *11*(1), 1-33. https://doi.org/10.1145/3470742

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*, 114-126. http://dx.doi.org/10.1037/xge0000033

Edgington, E. S. (1996). Randomized single-subject experimental designs. *Behaviour Research and Therapy, 34*, 567-574. https://doi.org/10.1016/00057967(96)00012-5

Elliott, M. N., McCaffrey, D. F., & Lockwood, J. R. (2007). How important is exact balance in treatment and control sample sizes to evaluations? *Journal of Substance Abuse Treatment, 33*(1), 107-110. https://doi.org/10.1016/j.jsat.2006.12.007

Eyssel, F., & Hegel, F. (2012). (S)he’s got the look: Gender stereotyping of robots. *Journal of Applied Social Psychology*, *42*(9), 2213-2230. https://doi.org/10.1111/j.1559-1816.2012.00937.x

Fallatah, A., Urann, J., & Knight, H. (2019, November 3-8). The robot show must go on: Effective responses to robot failures. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 325-332). IEEE. https://doi.org/10.1109/IROS40897.2019.8967854

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191. https://doi.org/10.3758/BF03193146

Ferrari, F., Paladino, M. P., & Jetten, J. (2016). Blurring human–machine distinctions: Anthropomorphic appearance in social robots as a threat to human distinctiveness. *International Journal of Social Robotics*, *8*(2), 287-302. https://doi.org/10.1007/s12369-016-0338-y

Ferron, J. M., Moeyaert, M., Van den Noortgate, W., & Beretvas, S. N. (2014). Estimating causal effects from multiple-baseline studies: Implications for design and analysis. *Psychological Methods, 19*(4), 493-510. https://doi.org/10.1037/a0037038

Fraune, M. R., Šabanović, S., & Smith, E. R. (2017, August 28-September 1). Teammates first: Favoring ingroup robots over outgroup humans. In *26th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 1432-1427). IEEE. https://doi.org/10.1109/ROMAN.2017.8172492

Giner-Sorolla, R. (2018, January 24). Powering your interaction. *Approaching significance.* https://approachingblog.wordpress.com/2018/01/24/powering-your-interaction-2/

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

Gompei, T., & Umemuro, H. (2015, August 31-September 4). A robot's slip of the tongue: Effect of speech error on the familiarity of a humanoid robot. In *24th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 331-336). IEEE. https://doi.org/10.1109/ROMAN.2015.7333630

Gonsior, B., Sosnowski, S., Mayer, C., Blume, J., Radig, B., Wollherr, D., & Kühnlenz, K. (2011, July 31- August 3). Improving aspects of empathy and subjective performance for HRI through mirroring facial expressions. In *20th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 350-356). IEEE. https://doi.org/10.1109/ROMAN.2011.6005294

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812), 619. https://doi.org/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. https://doi.org/10.1016/j.cognition.2012.06.007

Grover, S. L., & Brockner, J. (1989). Empathy and the relationship between attitudinal similarity and attraction. *Journal of Research in Personality*, *23*(4), 469-479. https://doi.org/10.1016/0092-6566(89)90015-9

Grundke, A., Stein, J.-P., & Appel, M. (2022). Mind-reading machines: Distinct user responses to thought-detecting and emotion-detecting robots. *Technology, Mind, and Behavior*, *3*(1), 1-12. https://doi.org/10.1037/tmb0000053

Hanson, D. (2005, December 5). Expanding the aesthetic possibilities for humanoid robots. In *IEEE-RAS International Conference on Humanoid Robots* (pp. 24-31). Citeseer. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.472.2518&rep=rep1&type=pdf

Hasson, Y., Tamir, M., Brahms, K. S., Cohrs, J. C., & Halperin, E. (2018). Are liberals and conservatives equally motivated to feel empathy toward others? *Personality and Social Psychology Bulletin*, *44*(10), 1449-1459. https://doi.org/10.1177/0146167218769867

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). The Guilford Press.

Heinke, M. S., & Louis, W. R. (2009). Cultural background and individualistic–collectivistic values in relation to similarity, perspective taking, and empathy. *Journal of Applied Social Psychology, 39*(11), 2570-2590. https://doi.org/10.1111/j.1559-1816.2009.00538.x

Hernández-Méndez, J., & Muñoz-Leiva, F. (2015). What type of online advertising is most effective for eTourism 2.0? An eye tracking study based on the characteristics of tourists. *Computers in Human Behavior, 50*, 618-625. https://doi.org/10.1016/j.chb.2015.03.017

Hildt, E. (2019). Artificial intelligence: Does consciousness matter? *Frontiers in Psychology*, *10*, Article 1535. https://doi.org/10.3389/fpsyg.2019.01535

Hoffmann, E. T. A. (2012). *Der Sandmann* (R. Drux, Ed.). Reclam Verlag. (Original work published 1816)

Horstmann, A. C., & Krämer, N. C. (2019). Great expectations? Relation of previous experiences with social robots in real life or in the media and expectancies based on qualitative and quantitative assessment. *Frontiers in Psychology, 10*, Article 939. https://doi.org/10.3389/fpsyg.2019.00939

Horstmann, A. C., Bock, N., Linhuber, E., Szczuka, J. M., Straßmann, C., & Krämer, N. C. (2018). Do a robot's social skills and its objection discourage interactants from switching the robot off? *PloS one*, *13*(7), Article e0201581. https://doi.org/10.1371/journal.pone.0201581

International Federation of Robotics. (2021). *Executive Summary World Robotics 2021 - Service Robots.* https://ifr.org/img/worldrobotics/Executive_Summary_WR_Service_Robots_2021.pdf

Kahn, P. H., Jr., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., Ruckert, J. H., & Shen, S. (2012). "Robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Developmental Psychology, 48*(2), 303-314. https://doi.org/10.1037/a0027033

Kang, J., & Sundar, S. S. (2019, June 26-28). Social robots with a theory of mind (ToM): Are we threatened when they can read our emotions? In *International Symposium on Ambient Intelligence* (pp. 80-88). Springer. https://doi.org/10.1007/978-3-030-24097-4_10

Kaseweter, K. A., Drwecki, B. B., & Prkachin, K. M. (2012). Racial differences in pain treatment and empathy in a Canadian sample. *Pain Research and Management*, *17*(6), 381-384. https://doi.org/10.1155/2012/803474

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P. D., Jewell, R., & Winter, N. J. (2020). The shape of and solutions to the MTurk quality crisis. *Political Science Research and Methods, 8*(4), 614-629. https://doi.org/10.1017/psrm.2020.6

Klein, K. J. K., & Hodges, S. D. (2001). Gender differences, motivation, and empathic accuracy: When it pays to understand. *Personality Social Psychology Bulletin, 27,* 720-730. http://dx.doi.org/10.1177/0146167201276007

Koć-Januchta, M., Höffler, T., Thoma, G. B., Prechtl, H., & Leutner, D. (2017). Visualizers versus verbalizers: Effects of cognitive style on learning with texts and pictures–An eye-tracking study. *Computers in Human Behavior*, *68*, 170-179. https://doi.org/10.1016/j.chb.2016.11.028

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods, 15,* 124-144. https://doi.org/10.1037/a0017736

Küster, D., & Swiderska, A. (2021). Seeing the mind of robots: Harm augments mind perception but benevolent intentions reduce dehumanisation of artificial entities in visual vignettes. *International Journal of Psychology, 56*(3), 454-465. https://doi.org/10.1002/ijop.12715

Lachin, J. M. (1988). Properties of simple randomization in clinical trials. *Controlled Clinical Trials, 9*(4), 312-326. https://doi.org/10.1016/0197-2456(88)90046-3

Laird, J. E., Lebiere, C., & Rosenbloom, P. S. (2017). A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *AI Magazine*, *38*(4), 13-26. https://doi.org/10.1609/aimag.v38i4.2744

Lee, K. M., Jung, Y., Kim, J., & Kim, S. R. (2006). Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *International Journal of Human-Computer Studies, 64*(10), 962-973. https://doi.org/10.1016/j.ijhcs.2006.05.002

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

Liu, Y., & Liao, S. (2021, June 18-20). Would humans want to work side-by-side with autonomous robots? The effect of robot autonomy on perceived usefulness, ease of use and desire for contact. In *Proceedings of the 2021 International Conference on Control and Intelligent Robotics* (pp. 671-675). ACM. https://doi.org/10.1145/3473714.3473830

Lotz-Schmitt, K., Siem, B., & Stürmer, S. (2017). Empathy as a motivator of dyadic helping across group boundaries: The dis-inhibiting effect of the recipient's perceived benevolence. *Group Processes & Intergroup Relations*, *20*(2), 233-259. https://doi.org/10.1177/1368430215612218

Macaskill, A., Maltby, J., & Day, L. (2002). Forgiveness of self and others and emotional empathy. *Journal of Social Psychology, 142*, 663-665. http://dx.doi.org/10.1080/00224540209603925

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

MacDorman, K. F., Green, R. D., Ho, C. C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior*, *25*(3), 695-710. https://doi.org/10.1016/j.chb.2008.12.026

Malinowska, J. K. (2021). What does it mean to empathise with a robot? *Minds and Machines, 31*, 361-376. https://doi.org/10.1007/s11023-021-09558-7

Mara, M., Appel, M., & Gnambs, T. (2022). Human-like robots and the uncanny valley: A meta-analysis of user responses based on the Godspeed Scales. *Zeitschrift für Psychologie*, *230*(1), 33-46. https://doi.org/10.1027/2151-2604/a000486

Mara, M., Stein, J. P., Latoschik, M. E., Lugrin, B., Schreiner, C., Hostettler, R., & Appel, M. (2021). User responses to a humanoid robot observed in real life, virtual reality, 3D and 2D. *Frontiers in Psychology, 12*, Article 633178. https://doi.org/10.3389/fpsyg.2021.633178

Mathur, M. B., & Reichling, D. B. (2009, March 9-13). An uncanny game of trust: Social trustworthiness of robots inferred from subtle anthropomorphic facial cues. In *4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 313-314). IEEE. https://doi.org/10.1145/1514095.1514192

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

Mattiassi, A. D., Sarrica, M., Cavallo, F., & Fortunati, L. (2021). What do humans feel with mistreated humans, animals, robots, and objects? Exploring the role of cognitive empathy. *Motivation and Emotion, 45*, 543-555. https://doi.org/10.1007/s11031-021-09886-2

McAuliffe, W. H., Carter, E. C., Berhane, J., Snihur, A. C., & McCullough, M. E. (2020). Is empathy the default response to suffering? A meta-analytic evaluation of perspective taking's effect on empathic concern. *Personality and Social Psychology Review, 24*(2), 141-162. https://doi.org/10.1177/1088868319887599

Mehrabian, A., Young, A. L., & Sato, S. (1988). Emotional empathy and associated individual differences. *Current Psychology*, *7*(3), 221-240. http://dx.doi.org/10.1007/BF02686670

Menne, I. M., & Schwab, F. (2018). Faces of emotion: Investigating emotional facial expressions towards a robot. *International Journal of Social Robotics, 10*(2), 199-209. https://doi.org/10.1007/s12369-017-0447-2

Meuwese, R., Cillessen, A. H., & Güroğlu, B. (2017). Friends in high places: A dyadic perspective on peer status as predictor of friendship quality and the mediating role of empathy and prosocial behavior. *Social Development*, *26*(3), 503-519. https://doi.org/10.1111/sode.12213

Michalos, G., Makris, S., Tsarouchi, P., Guasch, T., Kontovrakis, D., & Chryssolouris, G. (2015). Design considerations for safe human-robot collaborative workplaces. *Procedia CirP*, *37*, 248-253. https://doi.org/10.1016/j.procir.2015.08.014

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

Mittring, G. (2004). *Die Ermittlung der kleinsten hinreichend großen Stichprobe bei wissenschaftlichen Experimenten mit Randomisierung* [Doctoral dissertation, University of Cologne]. Networked Digital Library of Theses and Dissertations.

Moon, Y., & Nass, C. (1996). How “real” are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication Research*, *23*(6), 651-674. https://doi.org/10.1177/009365096023006002

Mori, M. (1970). The uncanny valley. *Energy, 7*(4), 33-35.

Müller, B. C., Gao, X., Nijssen, S. R., & Damen, T. G. (2020). I, Robot: How human appearance and mind attribution relate to the perceived danger of robots. *International Journal of Social Robotics*, *13*, 691-701. https://doi.org/10.1007/s12369-020-00663-8

Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied*, *7*(3), 171-81. https://doi.org/10.1037/1076-898X.7.3.171

Nass, C., & Moon, Y. (2002). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81-103. http://dx.doi.org/10.1111/0022-4537.00153

Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology*, *27*(10), 864-876. https://doi.org/10.1111/j.1559-1816.1997.tb00275.x

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353-383. https://doi.org/10.1016/0010-0285(77)90012-3

Nijssen, S. R., Müller, B. C., Baaren, R. B. V., & Paulus, M. (2019). Saving the robot or the human? Robots who feel deserve moral care. *Social Cognition*, *37*(1), 41-56. https://doi.org/10.1521/soco.2019.37.1.41

Osborne-Crowley, K., Wilson, E., De Blasio, F., Wearne, T., Rushby, J., & McDonald, S. (2019). Empathy for people with similar experiences: Can the perception-action model explain empathy impairments after traumatic brain injury? *Journal of Clinical and Experimental Neuropsychology, 42*(1), 28-41. https://doi.org/10.1080/13803395.2019.1662375

Oswald, P. A. (1996). The effects of cognitive and affective perspective taking on empathic concern and altruistic helping. *The Journal of Social Psychology*, *136*(5), 613-623. https://doi.org/10.1080/00224545.1996.9714045

Otterbacher, J., & Talias, M. (2017, March 6-9). S/he's too warm/agentic! The influence of gender on uncanny reactions to robots. In *12th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 214-223). IEEE. https://ieeexplore.ieee.org/document/8534886

Palomäki, J., Kunnari, A., Drosinou, M., Koverola, M., Lehtonen, N., Halonen, J., Repo, M., & Laakasuo, M. (2018). Evaluating the replicability of the uncanny valley effect. *Heliyon*, *4*(11), Article e00939. https://doi.org/10.1016/j.heliyon.2018.e00939

Perugini, M., Gallucci, M., & Costantini, G. (2014). Safeguard power as a protection against imprecise power estimates. *Perspectives on Psychological Science*, *9*(3), 319-332. https://doi.org/10.1177/1745691614528519

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*(9), 998-1000. https://doi.org/10.1068/p7569

Ragni, M., Rudenko, A., Kuhnert, B., & Arras, K. O. (2016, August 26-31). Errare humanum est: Erroneous robots in human-robot interaction. In *25th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 501-506). IEEE. https://doi.org/10.1109/ROMAN.2016.7745164

Reeves, B., & Nass, C. (1996). *The media equation–How people treat computers, television, and new media like real people and places.* CSLI Publications.

Robinette, P., Howard, A., Wagner, A.R. (2017). Conceptualizing overtrust in robots: Why do people trust a robot that previously failed? In W. Lawless, R. Mittu, D. Sofge, & S. Russell (Eds.), *Autonomy and artificial intelligence: A threat or savior?* Springer. https://doi.org/10.1007/978-3-319-59719-5_6

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. https://doi.org/10.1016/j.chb.2014.03.066

Rosenthal-von der Pütten, A.M., Krämer, N.C., Hoffmann, L., Sobieraj, S. & Eimler, S.C. (2013). An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics*, *5*, 17-34. https://doi.org/org/10.1007/s12369-012-0173-8

Rueckert, L., & Naybar, N. (2008). Gender differences in empathy: The role of the right hemisphere. *Brain and Cognition, 67,* 162-167. http://dx.doi.org/10.1016/j.bandc.2008.01.002

Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., & Joublin, F. (2013). To err is human(-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics*, *5*(3), 313-323. https://doi.org/10.1007/s12369-013-0196-9

Seo, S. H., Geiskkovitch, D., Nakane, M., King, C., & Young, J. E. (2015, March 2-5). Poor thing! Would you feel sorry for a simulated robot? A comparison of empathy toward a physical and a simulated robot. In *10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 125-132). IEEE. http://dx.doi.org/10.1145/2696454.2696471

Simonsohn, U. (2014, March 12). [17] No-way interactions. *Data Colada.* http://datacolada.org/17

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, *1156*(1), 81-96. https://doi.org/10.1111/j.1749-6632.2009.04418.x

Stapels, J. G., & Eyssel, F. (2022). Robocalypse? Yes, please! The role of robot autonomy in the development of ambivalent attitudes towards robots. *International Journal of Social Robotics*, *14*, 683-697. https://doi.org/10.1007/s12369-021-00817-2

Steain, A., Stanton, C. J., & Stevens, C. J. (2019). The black sheep effect: The case of the deviant ingroup robot. *PloS one*, *14*(10), Article e0222975. https://doi.org/10.1371/journal.pone.0222975

Stein, J. P., Appel, M., Jost, A., & Ohler, P. (2020). Matter over mind? How the acceptance of digital entities depends on their appearance, mental prowess, and the interaction between both. *International Journal of Human-Computer Studies*, *142*, Article 102463. https://doi.org/10.1016/j.ijhcs.2020.102463

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

Taylor, J., Weiss, S. M., & Marshall, P. J. (2020). "Alexa, how are you feeling today?": Mind perception, smart speakers, and uncanniness. *Interaction Studies*, *21*(3), 329-352. https://doi.org/10.1075/is.19015.tay

van den Brule, R., Dotsch, R., Bijlstra, G., Wigboldus, D. H., & Haselager, P. (2014). Do robot performance and behavioral style affect human trust? *International Journal of Social Robotics*, *6*(4), 519-531. https://doi.org/10.1007/s12369-014-0231-5

Vanman, E. J., & Kappas, A. (2019). "Danger, Will Robinson!" The challenges of social robots for intergroup relations. *Social and Personality Psychology Compass*, *13*(8), Article e12489. https://doi.org/10.1111/spc3.12489

Ward, A. F., Olsen, A. S., & Wegner, D. M. (2013). The harm-made mind: Observing victimization augments attribution of minds to vegetative patients, robots, and the dead. *Psychological Science, 24*(8), 1437-1445. https://doi.org/10.1177/0956797612472343

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology, 52*, 113-117. http://dx.doi.org/10.1016/j.jesp.2014.01.005

Yam, K. C., Bigman, Y. E., Tang, P. M., Ilies, R., De Cremer, D., Soh, H., & Gray, K. (2020). Robots at work: People prefer—and forgive—service robots with perceived feelings. *Journal of Applied Psychology, 106*(10), 1557-1572. http://dx.doi.org/10.1037/apl0000834

Yang, G. Z., J. Nelson, B., Murphy, R. R., Choset, H., Christensen, H., H. Collins, S., Dario, P., Goldberg, K., Ikuta, K., Jacobstein, N., Kragic, A., Taylor, R.H., & McNutt, M. (2020). Combating COVID-19—The role of robotics in managing public health and infectious diseases. *Science Robotics*, *5*(40), Article eabb5589. https://doi.org/10.1126/scirobotics.abb5589

Young, A. D., & Monroe, A. E. (2019). Autonomous morals: Inferences of mind predict acceptance of AI behavior in sacrificial moral dilemmas. *Journal of Experimental Social Psychology, 85*, Article 103870. https://doi.org/10.1016/j.jesp.2019.103870

Young, A., Khalil, K. A., & Wharton, J. (2018). Empathy for animals: A review of the existing literature. *Curator: The Museum Journal*, *61*(2), 327-343. https://doi.org/10.1111/cura.12257

Zafari, S., & Koeszegi, S. T. (2020). Attitudes toward attributed agency: Role of perceived control. *International Journal of Social Robotics*, *13*(8), 2071-2080. https://doi.org/10.1007/s12369-020-00672-7