Check for updates

Article

new media & society
1–20
© The Author(s) 2020

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1461444820954199
journals.sagepub.com/home/nms


I saw it on YouTube! How online videos shape perceptions of mind, morality, and fears about robots

Dennis Küster
University of Bremen, Germany

Aleksandra Swiderska
University of Warsaw, Poland

David Gunkel
Northern Illinois University, USA

Abstract

Robots have the potential to transform our existing categorical distinctions between "property" and "persons." Previous research has demonstrated that humans naturally anthropomorphize them, and this tendency may be amplified when a robot is subject to abuse. Simultaneously, robots give rise to hopes and fears about the future and our place in it. However, most available evidence on these mechanisms is either anecdotal, or based on a small number of laboratory studies with limited ecological validity. The present work aims to bridge this gap through examining responses of participants ($N = 160$) to four popular online videos of a leading robotics company (Boston Dynamics) and one more familiar vacuum cleaning robot (Roomba). Our results suggest that unexpectedly human-like abilities might provide more potent cues to mind perception than appearance, whereas appearance may attract more compassion and protection. Exposure to advanced robots significantly influences attitudes toward future artificial intelligence. We discuss the need for more research examining groundbreaking robotics outside the laboratory.

Corresponding author:
Dennis Küster, Department of Computer Science, University of Bremen, Enrique-Schmidt-Str. 5, 28359 Bremen, Germany.
Email: kuester@uni-bremen.de

Keywords
Agency, dehumanization, harmfulness, moral typecasting, robots

Introduction

In confronting and dealing with other entities—whether other human persons, non-human animals, or technological artifacts—one inevitably needs to distinguish between those beings *who* are recognized as another social subject and *what* remains a mere thing or instrument. As the French philosopher Jacques Derrida (2005: 80) explained, the difference between these two small and seemingly insignificant words—"who" and "what"—makes a difference, precisely because it parses the world of entities into two camps: Those *who* count as another socially significant subject and *what* are and remain mere objects. Or, if one prefers to employ legal terminology, there is a difference between "persons" and "property." Robots and other forms of socially interactive technology challenge this categorical distinction in ways that often blur or complicate the seemingly simple binary distinction between "who" and "what." Consider two recent examples of hotly debated anecdotes and stories:

1. Hitchbot—Hitchbot (http://www.hitchbot.me/) was a hitchhiking robot, created by David Harris Smith and Frauke Zeller, which successfully made its way across Canada and Europe only to be brutally vandalized in August 2015 at the beginning of a similar effort to cross the United States. What is remarkable about the demise of Hitchbot is less the fact that it was vandalized. What is more noteworthy is the way human beings responded to this event. "Honestly," Kate Darling (2015) explains, "I was a little surprised that it took this long for something bad to happen to Hitchbot. But I was even more surprised by the amount of attention that this case got. I mean it made international headlines and there was an outpouring of sympathy and support from thousands and thousands of people for Hitchbot." To support her point, Darling referred to tweets from a large number of individuals who not only expressed a sense of loss over the "death of Hitchbot"—but who apologized directly to the robot for the cruelty done to it by humans (Gunkel, 2018).
2. Kicking a Robot—In the same year (2015), Boston Dynamics debuted a number of video recordings designed to demonstrate the unique balancing capabilities and dexterity of their robots, specifically the bipedal robot Atlas and the dog-like Spot. As part of this demonstration, human beings were pictured striking and pushing Atlas with a hockey stick and kicking the spot robot as it walked between office cubicles. Viewers of the videos, which quickly went viral via distribution on YouTube, were outraged. As CNN reported (https://www.cnn.com/2015/02/13/tech/spot-robot-dog-google/index.html), Twitter immediately lit up with comments:

> Seriously Boston Dynamics stop kicking those poor robots what did they ever do to you?!—@therealcliffyb (10 February 2015)

> Robots are so cool. I actually felt sorry for this one when it was kicked. It looked so dog-like. Poor Spot!—@RZKelleher (11 February 2015)

The outcry was so wide-ranging and pervasive, that PETA (People for the Ethical Treatment of Animals) found it necessary to get involved:

> PETA deals with actual animal abuse every day, so we won't lose sleep over this incident. But while it's far better to kick a four-legged robot than a real dog, most reasonable people find even the idea of such violence inappropriate, as the comments show. (Parke, 2015)

These two events provide anecdotal evidence that the way we typically make sense of and divide up the world of entities is straining against existing categorical distinctions. Furthermore, they appear to point to broader trends in how robots and artificial intelligence (AI) elicit hopes and fears in response to innovations in this field. Some of these devices have already reached our households, as evidenced by the now familiar and almost ubiquitous presence of vacuum cleaner and lawn mowing robots. Nevertheless, most people still lack firsthand experience with the kind of advanced robots developed by Boston Dynamics. Thus, there is both constant, slow pressure on existing categorical distinctions, and the potential for more rapid re-mappings of how we perceive "robotic others" when people are exposed to more sophisticated capabilities and the appearance of such entities.

Although there has been some effort to investigate these matters in experimental studies, much more is needed to understand not just the artifact that is the robot but, more importantly, the way we decide to respond to and take responsibility for what happens in the face of these devices. More specifically, there is still a wide gap between the type of powerful yet anecdotal discussion about "futuristic" robots and their place in human society, and the limited amount of highly controlled experimental studies in which human subjects have been exposed to depictions of comparatively primitive robotic entities. We therefore posed the following research questions: To what extent is public perception of robots today still a matter of (or a lack of) everyday experience? Does viewing a regular testing session of advanced robot capabilities suffice to instill a sense of the "other" as a *who* with a mind worthy of protection from harm—or does this only occur in anecdotes or industry demonstrations that were explicitly designed for this purpose? Our study, which documents and examines the responses of participants ($N = 160$) to video-recorded illustrations of robot abuse, seeks to provide empirically tested evidence for addressing and resolving these questions.

Overview of prior research

Behavioral responses to artificial entities: Media Equation and robot abuse

People tend to accord social standing and personal respect to technological objects. This insight was initially demonstrated and evaluated in "An Experimental Study of Apparent Behavior" by Fritz Heider and Marianne Simmel (1944). They found that human subjects attributed motives and personality to simple animated geometric figures. Similar

results were obtained and validated in the "Computers As Social Actors" (CASA) studies, conducted by Byron Reeves and Clifford Nass in the mid-1990s. In essence, as demonstrated in several studies (e.g. Nass and Moon, 2000; Nass et al., 1997; Reeves and Nass, 1996), users often respond to socially interactive technology as if they were other people—even when this technology is only basic or rudimentary. For example, study participants adhered to politeness norms when responding to computers or engaged in reciprocal self-disclosure (Moon, 2000; Nass et al., 1999). These effects were labeled the Media Equation (i.e. media equals real life; Reeves and Nass, 1996) and explained to originate in mindless execution of behavioral scripts of Human–Human Interactions (HHI) in interactions with computers (Nass and Moon, 2000).

The Media Equation has been extended to robots by way of a research paradigm that is now called "robot abuse" studies (Bartneck et al., 2005; Brahnam and De Angeli, 2008; De Angeli et al., 2005, 2006). This paradigm comprises what has been characterized as "the dark side" of research in Human–Robot Interaction (HRI; De Angeli et al., 2005). As Bartneck and Hu (2008) explain, the main reason for this kind of work has been to examine the limits of HRI by intentionally stepping outside the boundaries of normal conduct:

> It is only from an extreme position that the applicability of the Media Equation to robots might become clear. In our study we have therefore focused on robot abuse. What we propose to investigate in this context is whether human beings abuse robots in the same way as they abuse other human beings, as suggested by the Media Equation. (p. 416)

These robot abuse studies have contributed to two different but related research vectors. First, they experimentally tested the accuracy of the Media Equation as it applies (or not) to robots and other autonomous devices. Bartneck et al. (2008), for example, have designed and performed Milgram-like obedience experiments, finding that human subjects are less inhibited when instructed to administer harm to robots than to human beings. This outcome not only challenges the predictions of the Media Equation but has been used as evidence of significant limitations regarding the applicability of the theory to robotic artifacts. Second, robot abuse studies are motivated by the needs of device design strategies and implications. As described by De Angeli et al. (2006), "the overarching objective [of this work] . . . is to sketch a research agenda on the topic of the misuse and abuse of interactive technologies that will lead to design solutions capable of protecting users and restraining disinhibited behaviors." (p. 1) Existing robot abuse studies, therefore, have been conducted for design testing and are not necessarily pursued as the means for investigating human expectations for robots and the perception of social standing and value.

Emotional responses and empathy toward robots

In more recent work, there has been some effort to extend and develop the "robot abuse" research paradigm in the direction of user perceptions and social effects. Indicative of this development is a study conducted by Rosenthal-von der Pütten et al. (2013), where researchers investigated the emotional reaction of users to the viewing of two video

recordings, one representing "robot torture" and the other consisting of "friendly interactions." The study found that participants responded with elevated physiological arousal to the torture video compared with the friendly video, and "expressed empathic concern for the robot" (p. 17). These results contravene the finding of Bartneck et al. (2008), lending credibility to the predictions of the Media Equation and providing data supporting "assumptions on the socialness of reactions towards robots and anecdotal evidence of emotional attachments to robots." In another study, Suzuki et al. (2015) used electroencephalography (EEG) to measure the response of human subjects to human and robot harm as represented by images of painful events, such as a knife cutting into a finger. The results indicated that participants empathized with the pain of the robot in ways that were similar to that of another human being.

Perception of mind and moral status

The discrepancies in responses to robots in the context of The Media Equation, and the differences in the inhibition to abuse them, point to a host of questions about the underlying psychological mechanisms. When do we begin to perceive robots as a "who" worthy of consideration, and when are they just a mere object that can be carelessly destroyed? Prior research demonstrated that social cues, such as human-like appearance and behavioral realism, lead the human perceiver to anthropomorphize artificial entities, that is, to imbue them with a human-like mind (Epley et al., 2007; Von der Pütten et al., 2010). It has been further found that others' minds are perceived along two separate dimensions: *experience* and *agency* (Gray et al., 2007). Experience is an umbrella term for a variety of capacities related to sensing and feeling (e.g. feelings of pleasure, pain, or hunger—but also consciousness and personality), while agency encompasses capacities linked to planning and acting (e.g. self-control, memory, and communicative abilities; Gray et al., 2007; Waytz et al., 2010). Intriguingly, experience has been found to reveal more pronounced distinctions in mind attributions toward humans and robots than agency. Thus, robots tend to be perceived as further away from a human-level capacity for experience than for attributions of agency (Gray et al., 2007, 2012). In this framework, mind perception is not an all-or-none phenomenon, but rather a matter of degree, detached from whether there exists an actual biological basis for an entity to possess mental states (e.g. Gray et al., 2011).

Out of different technological devices, robots in particular benefit from both anthropomorphism and zoomorphism, and hence the increasing tendency in the field of robotics to build human-like or animal-like machines (Bartneck et al., 2006; Złotowski et al., 2015). Their physical presence in the real world broadens the scope of possible social reactions toward them, for example, when they are regarded to be more accountable for their behavior than only virtually present computer agents (Kahn et al., 2012). Overall, anthropomorphic and zoomorphic qualities have been argued to render interactions with robots more natural and understandable for the user (e.g. Breazeal, 2002; cf. Jones et al., 2008). Sometimes, however, a high degree of human likeness can actually be detrimental to the interaction and cause the user to like and relate to the robot less due to difficulties with distinguishing it from a living human (Mori et al., 2012). As suggested by recent evidence, a similar mechanism might hold also for zoomorphic animals (Löffler et al.,

2020). This may be especially the case when the robot appears to have an increased capacity for experience (Gray and Wegner, 2012). In line with this finding, another study has shown that people readily attribute the capacity for agency to robots, but less so the capacity for experience (Gray et al., 2007). Interestingly, having emotions and personality (i.e. key aspects of experience) are readily attributed to dogs (Konok et al., 2018). This has raised the question of how such nonhuman animals could inspire the design of future zoomorphic robots (Konok et al., 2018), how human–animal interaction might inform building animal-like robotic pets (Miklósi and Gácsi, 2012), and how animal-like robots might help to better understand which robot features and behaviors are specifically tied to the perception of human capabilities (Jones and Schmidlin, 2011).

In psychological literature, how people attribute experience and agency to others is intertwined with how they construe moral interactions. An example of a moral interaction is a situation in which one party (an agent) inflicts harm on another (a patient), like in the robot abuse studies. It has been shown that for entities of low or even nonexistent level of mental capacities, such as a person in a vegetative state or a corpse, mind perception increased when these entities were subject to intentional abuse (Ward et al., 2013). This effect—*the harm-made mind*—occurs for robots as well (Swiderska and Küster, 2018; Ward et al., 2013). Toward an explanation, the mechanism of an automatic completion of a moral dyad was invoked. That is, if there is a (moral agent's) mind capable of conceiving and carrying out a harmful action, there has to be a second independent mind to experience the outcomes of this action (e.g. via feeling pain; Gray and Wegner, 2009; Ward et al., 2013). The ascription of these roles in a moral interaction is in turn associated with moral judgment—malevolent agents are seen to be deserving of punishment, while patients are entitled to protection (Gray et al., 2007).

Overview of the present research

Our study aimed to better identify, document, and evaluate human perception of robot harm and social status, as well as general attitudes toward intelligent robots. We established a fine-grained evaluation mechanism for measuring the impact and significance of these perceptions, and made the following predictions: First, in line with prior work on the harm-made mind effect, we hypothesized that abused robots would be attributed mental capacities (i.e. pain, experience, consciousness, and agency) to a higher extent compared with their non-abused counterparts. We expected this effect to be more pronounced for a more human-like robot (Boston Dynamics' Atlas) compared with a mechanical, animal-like robot (Boston Dynamics' Spot) since human-like appearance should facilitate mind perception. In the same vein, we hypothesized the human-like robot to be granted more rights associated with being a moral patient than the animal-like robot as the assignment of moral worth should be paired with mind perception. In line with the uncanny valley theory (Mori et al., 2012), we further hypothesized the human-like robot to elicit the highest degree of discomfort, compared with the animal-like robot, as well as compared with an additional nonhuman-like control condition: a Roomba vacuum cleaning robot.

Finally, we aimed to examine to what extent exposure to highly advanced "cutting edge" robots on the Internet (e.g. via YouTube) might shape more general attitudes

toward AI. While recent work suggests a sharp increase in, often optimistic, discussion about AI in recent decades, worries about negative impacts, ethical concerns, and loss of (human) control appear to be on the rise (Fast and Horvitz, 2017). Here, we expected exposure to the Boston Dynamics robots to elevate hopes and fears about AI, relative to watching the already more familiar Roomba. From a more general perspective, our approach thus broadens the view from individual responses to a harmed artificial entity in the lab toward a consideration of the societal impact of impending leaps in development of highly human-like robot capabilities. We deem this aspect important because these kinds of current real-world examples have the potential to shape public perception and societal discourse in ways that more controlled yet fictitious experimental materials have been lacking.

Method

Participants

An a priori power analysis with G*Power (Faul et al., 2007) indicated that 172 participants[1] (43 per condition) would be sufficient to detect a medium-sized main effect (Cohen's $f = 0.25$, $\alpha = 0.05$) of our primary hypothesis on the main effects of harm, robot type (Atlas, Spot), and the corresponding interaction effect in a 2×2 analysis of variance (ANOVA) with 90% power. One-hundred sixty participants (108 women; $M_{age} = 35.02$ years, $SD = 12.42$) completed the experiment online. All of them were volunteers, recruited via Prolific (https://www.prolific.co/) and paid 1 GBP each for filling out an approximately 6-minute long survey (in English). The Ethics Committee at the Department of Psychology, University of Warsaw, Poland approved the study.

Materials

We used five videos in total. Two videos introduced two robots of Boston Dynamics, an engineering company devoted to the creation of highly advanced machines (https://www.bostondynamics.com/about): A four-legged, dog-like Spot and a humanoid Atlas. These clips showed examples of specific abilities of the robots, such as opening the door in the case of Spot. Another two videos depicted the same robots in a situation where they were being hurt by a human (e.g. hit with a hockey stick) while they attempted to perform a certain action.[2] An additional video of a small vacuum cleaner Roomba served as an instance of a household robot that we expected participants to already be more familiar with.[3] The materials varied between 0:46 and 2:41 minutes in length and were publicly available on YouTube at the time of data collection.

Procedure and design

First, participants were requested to watch one randomly chosen video. The video was embedded directly in a questionnaire, delivered through Qualtrics XM 2019. Afterward, participants evaluated the robots in the videos on a series of dimensions, as well as reported their general attitudes toward AI (see dependent measures). The experiment

followed a 2 (robot type: Spot, Atlas) by 2 (harm: harmed, unharmed) between-subjects factorial design.

Dependent measures. Mind perception was assessed with ratings of the robots' perceived capacity for *experience* (seven items: having personality, experiencing desire, feelings, emotions, pleasure, hunger, fear; Cronbach's $\alpha = .93$), and *agency* (seven items: planning and controlling actions, remembering events, understanding others, understanding right from wrong, influencing situations, communicating; $\alpha = .83$), in line with Gray et al. (2007). In addition, we assessed the robots' perceived capacity to feel *pain* (one item) and *consciousness* (two items: being conscious of oneself and the world around; $\alpha = .75$), following Ward et al. (2013), who treated it as a more explicit measure of the endorsement of an entity's mental status. *Moral patiency* was estimated with two questions about whether the robots should be treated with compassion and fairness and whether they deserve to be protected from harm (two items; $\alpha = .88$), based on Loughnan et al. (2013). The response scales for these two measures ranged from 1 = *strongly disagree* to 7 = *strongly disagree*. Moreover, for the videos that depicted harm done to the robots, participants evaluated how morally right or wrong the human's behavior appeared to be (1 = *definitely wrong*, 7 = *definitely right*). We further included a measure of the extent to which robots were associated with *discomfort* (six items: scary, strange, awkward, dangerous, awful, aggressive; $\alpha = .87$; response scale from 1 = *definitely not associated* to 9 = *definitely associated*) from Carpinella et al. (2017), to examine if the highly advanced Boston Dynamics robots might influence discomfort experienced toward robots in general. The participants' perspectives on the presence of AI in diverse real-world applications was determined with a five-item scale of hopes for its positive outcomes (making human work easier, improving students' learning, enabling new forms of transportation, enhancing health, bringing joy through entertainment; $\alpha = .80$) and a four-item scale pertinent to concerns about negative outcomes (displacing jobs, killing people, losing control over AI, AI lacking ethical reasoning; $\alpha = .71$; response scale from 1 = *strongly disagree* to 7 = *strongly disagree*), selected from Fast and Horvitz (2017). Finally, we asked if participants were personally familiar with the type of a robot they saw in the video (1 = *not at all familiar*, 7 = *very familiar*).

Results

The first part of our analyses concerned the impact of the robots' human likeness and harm on mind perception, moral patiency perception, and evaluations of moral value of harmful behaviors toward the robots. Here, we only considered the Boston Dynamics videos that featured the dog-like Spot robot and the humanoid Atlas. The vacuum cleaner Roomba was not being harmed in the video and thus served as a comparison only in the subsequent analyses of familiarity, discomfort with robots, and attitudes toward AI. In these analyses, harm as such was not expected to play a role.

Mind perception

A multivariate ANOVA with Robot Type (Spot, Atlas) and Harm (harmed, unharmed) as between-subjects factors was conducted on pain, experience, agency, and consciousness

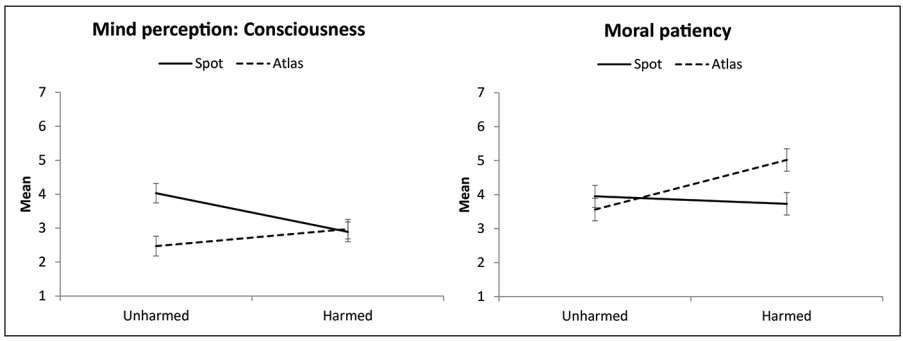

Figure 1. Evaluations of consciousness and moral patiency as a function of robot type and harm.

as the dependent variables,[4] with Bonferroni correction for multiple comparisons. The interaction effect between robot type and harm was significant, $F(4, 122) = 3.12$, $p = .018$, $\eta_p^2 = .09$, and so was the main effect of robot type, $F(4, 122) = 3.76$, $p = .006$, $\eta_p^2 = .11$.[5] The main effect of harm was not significant, $F(4, 122) = .62$, $p = .649$, $\eta_p^2 = .02$.

As revealed by the subsequent univariate tests, the interaction was significant only for consciousness, $F(1, 125) = 8.07$, $p = .005$, $\eta_p^2 = .06$. Spot was perceived as more conscious when it was not being harmed by a human, compared with when it was subject to abuse ($p = .006$), and unharmed Spot was also more viewed as more conscious than unharmed Atlas ($p < .001$; Figure 1). The difference in perceived consciousness between unharmed and harmed Atlas was not significant ($p = .224$), similarly to harmed Spot compared with harmed Atlas ($p = .854$). Thus, although participants attributed mental capacities to a higher degree to Spot than to Atlas, harm did not facilitate this process in the present context. The findings suggest that anthropomorphic appearance of an advanced robot does not necessarily correspond to increased mind perception, but rather, a demonstration of an unexpectedly human-like behavior, like the successful opening of a door by a dog-like entity, may be a more important cue to mind perception than appearance.

Patiency and moral evaluation

Next, we conducted an ANOVA for moral patiency. The interaction between robot type and harm was again significant, $F(1, 125) = 6.67$, $p = .011$, $\eta_p^2 = .05$. The main effect of robot type did not reach significance, $F(1, 125) = 1.91$, $p = .170$, $\eta_p^2 = .02$, and the main effect of harm was marginally significant, $F(1, 125) = 3.56$, $p = .061$, $\eta_p^2 = .03$. Harmed robots were granted slightly more moral patiency than unharmed robots, but most importantly, harmed Atlas was evaluated to deserve more compassion and protection than both unharmed Atlas ($p = .002$) and harmed Spot ($p = .006$; Figure 1). Here, the difference between unharmed and harmed Spot was not significant ($p = .623$), and the same was the case for the comparison between unharmed Spot and unharmed Atlas ($p = .399$).

The main effect of robot type was also significant for moral evaluation of the human agent's actions displayed in the videos, $F(1, 63) = 4.08$, $p = .048$, $\eta_p^2 = .06$. Harm toward Atlas seemed more morally wrong ($M_{Atlas} = 3.13$, $SD = 1.68$) than toward Spot ($M_{Spot} = 3.91$,

$SD = 1.44$). Nevertheless, when gender was included as a covariate ($p = .239$), this effect was only marginally significant, $F(1, 62) = 3.66$, $p = .060$, $\eta_p^2 = .06$.

The findings point to a possible dissociation between attributions of consciousness to an artificial entity, and the perceptions of both the entity's moral patiency and the severity of harmful behavior toward it. Thus, while the humanoid Atlas was perceived as less conscious than Spot, it still appears to have been viewed as more worthy of compassion and protection when harmed by a human engineer in the video, and harm toward it was seen as more morally wrong.

The remaining analyses focused on our hypothesis that the videos of highly advanced and possibly unfamiliar robots like Atlas and Spot should influence how naive participants perceive and think about robots and Artificial Intelligence (AI) in general. We expected participants to be more familiar with Roomba on the basis of their own personal experience. Moreover, we expected that Spot and Atlas evoke greater discomfort with robots than the familiar Roomba. These analyses were conducted with robot type as the independent variable. We collapsed across the harmed and unharmed conditions for Spot and Atlas, and included Roomba as a control condition.

Familiarity

First, we conducted an ANOVA with familiarity to examine if Roomba was indeed perceived as more familiar than Spot and Atlas. This analysis yielded a significant main effect of robot type, $F(2, 156) = 5.73$, $p = .004$, $\eta_p^2 = .07$.[6] Participants reported that they were least familiar with Atlas ($M = 2.41$, $SD = 1.65$) and, as expected, most familiar with Roomba ($M = 3.71$, $SD = 1.83$), and this difference was significant ($p = .003$). Spot fell in the middle ($M = 2.91$, $SD = 1.95$), with the difference between Spot and Roomba being marginally significant ($p = .059$). These results support the notion that the present sample of participants was rather unfamiliar with the advanced robots from Boston Dynamics, and that Roomba appeared substantially more familiar than Atlas, and somewhat more familiar than Spot.

Discomfort with robots

Our next aim was to examine whether watching the uncannily human-like capabilities of Spot and Atlas might lead to overall discomfort with robots. We further expected the humanoid Atlas to elicit the most discomfort. Exposure to the three types of videos influenced how much discomfort participants associated with robots overall, $F(2, 155) = 7.66$, $p = .001$, $\eta_p^2 = .09$.[7] However, contrary to our hypothesis, it was Spot that evoked the most discomforting associations ($M = 4.64$, $SD = 1.80$), followed by Atlas ($M = 4.22$, $SD = 1.79$). Importantly, both types of advanced Boston Dynamics robots elicited significantly stronger discomfort with robots than Roomba ($M = 3.13$, $SD = 1.99$; respectively, $p < .001$ and $p = .008$).

Attitudes toward AI

Finally, we investigated if the type of robot shown in the video impacted positive and negative attitudes toward AI. The multivariate main effect of robot type was significant,

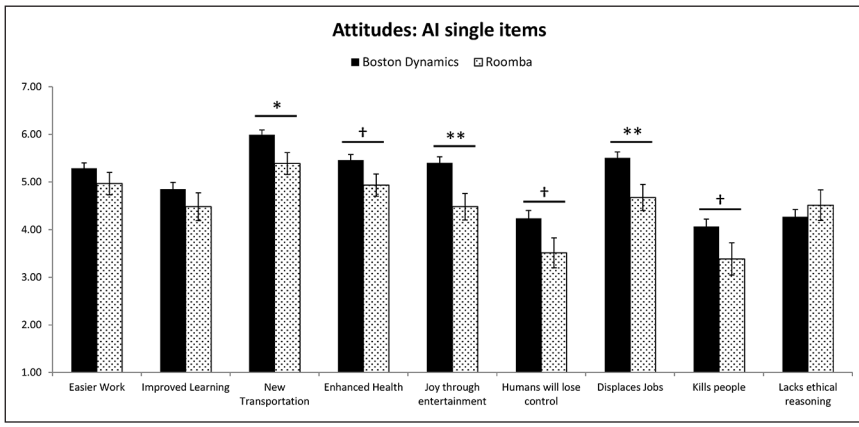

Figure 2. Effects of robot type on attitudes toward Artificial Intelligence (AI).
** $p < .01$; * $p < .05$; † $p < .10$.

$F(4, 312) = 3.38$, $p = .010$, $\eta_p^2 = .04$. Ensuing univariate tests demonstrated the main effect to be significant for hopes, $F(2, 157) = 4.17$, $p = .017$, $\eta_p^2 = .05$, and non-significant for concerns, $F(2, 157) = 3.35$, $p = .127$, $\eta_p^2 = .03$. In particular, participants were more optimistic about the presence of AI after the Spot video than after the Roomba video ($M_{Spot} = 5.49$, $SD = .89$ vs $M_{Roomba} = 4.85$, $SD = 1.04$, $p = .004$), even though the former was associated with more discomfort toward robots. Similarly, participants were more optimistic about AI after the Atlas video than after the Roomba video ($M_{Atlas} = 5.31$, $SD = 1.12$; $p = .042$).

The initial analyses supported our hypothesis that watching advanced robots influences attitudes toward AI. Nonetheless, since we did not obtain significant differences between Spot and Atlas, we proceeded to further collapse across all Boston Dynamics videos, and we examined the effects of exposure to them, compared with Roomba, at the level of nine individual AI-attitude statements (Figure 2).

The main effect of robot type (Boston Dynamics, Roomba) was significant, $F(9, 150) = 3.62$, $p < .001$, $\eta_p^2 = .18$. Regarding specific items, the differences were significant when it came to hopes for AI to enable new forms of transportation, $F(1, 158) = 6.62$, $p = .011$, $\eta_p^2 = .04$, and to bring people joy from entertainment, $F(1, 158) = 10.15$, $p = .002$, $\eta_p^2 = .06$, as well as marginally significant for AI to enhance people's health and well-being, $F(1, 158) = 3.76$, $p = .054$, $\eta_p^2 = .02$. The differences were also significant for concerns linked to AI displacing human jobs, $F(1, 158) = 8.00$, $p = .005$, $\eta_p^2 = .05$, and marginally significant for the possibility that people will lose control over powerful AI systems, $F(1, 158) = 3.81$, $p = .053$, $\eta_p^2 = .02$, and that AI could cause warfare and kill people through military applications, $F(1, 158) = 3.67$, $p = .057$, $\eta_p^2 = .02$. As demonstrated in Figure 2, the Boston Dynamics robots engendered higher ratings on all of these consequences of AI development than Roomba.

Overall, these results suggest that watching just a short film clip of highly advanced current robots on the Internet is likely to increase a relatively broad scope of individual hopes and fears about the future of Artificial Intelligence in our lives. However, as shown

by the results of the familiarity analysis, even the Roomba vacuum cleaning robot only achieved a moderate rating on a 7-point response scale. It is therefore possible that participants might have expressed even less intense hopes and concerns about AI if they had not watched any film clip, or a film clip featuring a more conventional vacuum cleaner.

Discussion

Function before form

From what we found, robot behavior appears to have a more significant impact on mind perception and anthropomorphic projection than morphology, and this insight is consistent with finding in recent human–machine communication (HMC) investigations (Banks, 2020; Edwards et al., 2016; Sandry, 2015). A robot does not necessarily need to look like a human for human research participants to attribute mind, motive, or internal states to it. Consequently, what the robot is shown to be capable of doing, that is, opening doors, is more determinative of mind perception than how it looks. Function appears to trump form.

This outcome, which follows from and contributes additional experimental evidence to the results initially reported by Heider and Simmel's (1944) "An Experimental Study of Apparent Behavior," has potentially important implications for the design of robots. If designers intend or desire to produce the effects of anthropomorphic projection in users of a robotic device, the morphology of the robot—whether it is humanoid in appearance, animal-like, caricatured with googly eyes, or entirely otherwise—may be less important than the behaviors and functions that the robot is able to deploy and display. The conditional form of this statement, however, is important. Although we have found that function takes precedence over form in producing the effects of anthropomorphic projection, it remains an open question whether and to what extent this effect is in fact a desirable objective. We were surprised by the intensity of responses elicited by Spot. However, in addition to being less human-like in shape, Spot's appearance and capabilities are also closely modeled after a type of pet that humans are well-known to react positively to, as dogs have co-evolved with and been bred by humans for millennia. Many people value dogs for their ability to show attachment and emotions, with dogs and other social animals being appreciated as a rich source for social robotics design (Konok et al., 2018; Miklósi and Gácsi, 2012). However, other related work on the role of behavior and/or appearance of crudely dog-like robots has suggested that individuals may vary substantially in their responses to such robots, demonstrating no advantage of either function or form (Jones et al., 2008). Therefore, it still stands to reason that for other kinds of devices, for example, insectoid robots, form and appearance will be more important. Based on what we know from anecdotal evidence concerning multi-legged military robots, we would expect that even completely alien-looking robots will trigger human empathy under the right circumstances. However, this does not mean that appearance would be irrelevant.

Patiency sans consciousness

Our results suggest that there is a likely dissociation between the attribution of consciousness and moral status. Whereas Spot was overall perceived to possess more experience,

agency and consciousness, harming Atlas appeared to be perceived as slightly more wrong than harming Spot, and the harmed Atlas robot was found to deserve more compassion and protection than a similarly harmed Spot. In other words, moral patiency—that is, the perception that the robot is a kind of socially situated other or a "who" as opposed to a mere instrument or a "what"—may not necessarily be predicated upon the assumed presence or the projected attribution of consciousness. Conversely, the perceived social and moral status of a robot— that is, the perception that it is worthy of compassion and protection when subjected to harmful actions by human users—seems to be at least partially decoupled from attributions of consciousness.

This outcome lends experimental credibility to the "relational ethics" introduced and developed by Mark Coeckelbergh (2012) and David Gunkel (2018). "The standard approach to the justification of moral status is," Coeckelbergh (2012) explains, "to refer to one or more (intrinsic) properties of the entity in question, such as consciousness . . . If the entity has this property, this then warrants giving the entity a certain moral status." (p. 13) According to this transaction, what something is determines how it ought to be treated. Or as Luciano Floridi (2013) describes it: "what the entity is determines the degree of moral value it enjoys, if any" (p. 116). Relational ethics flips the script on this procedure, proposing that questions regarding moral status, that is, how something is to be treated, are often independent of or disassociated with decisions (or assumptions) regarding the presence or absence of a qualifying property like consciousness. These results, however, are preliminary insofar as the investigation only found evidence of a possible dissociation between moral patiency and the attribution of consciousness. Additional work would be necessary to verify and test this conclusion for a broader range of different entities and contexts. However, if confirmed, this view would suggest that how we grant moral status to other entities is primarily determined by how it relates to us as human beings rather than by its apparent mental capabilities.

Proximity breeds discomfort

Results from our study seem to verify findings from previous investigations suggesting that the very mechanisms that allow for greater levels of human empathy with robots may also be those which are likely to produce the greatest levels of discomfort (Ceh and Vanman, 2018; Vanman and Kappas, 2019). This seemingly paradoxical outcome alludes to familiar notions harkening back to psychoanalysis, especially Sigmund Freud's concept of *das Unheimliche*, and the field of robotics, by way of Masahiro Mori's uncanny valley. Today, while there are still several competing explanations of the uncanny valley phenomenon (cf. MacDorman et al., 2009), one of the most influential accounts holds that this discomfort arises from conflicting cues at category boundaries and the resulting perceptual tension (Moore, 2012), for example, the seemingly alive dog that simultaneously still appears to be a robot. Although our study was not designed to investigate or test these hypotheses, our results appear generally aligned with theoretical predictions that proximity and discomfort are proportional. Thus, the familiar but unemotional vacuum cleaning robot elicited rather little discomfort in comparison with the Boston Dynamics robots, as well as fewer concerns. This finding is consistent with previous work suggesting that repeated interactions, or more generally familiarity with the

behaviors of a robot, are likely to reduce discomfort (Złotowski et al., 2018). However, our results also suggest that a highly human-like morphology is not necessarily what determines whether we perceive another entity as uncanny or discomforting.

We speculate that human-like morphology may not, as such, be uniquely associated with uncanny associations in response to a robot. Instead, our present findings appear to be consistent with the notion that zoomorphism may likewise result in discomfort comparable with the classic uncanny valley phenomenon (cf. Löffler et al., 2020). Spot's high ratings for experience and mind perception are in line with reports of dogs being particularly emotional and having a personality (Konok et al., 2018). As Konok et al. (2018) have suggested, dogs are often perceived as a lovable social partner capable of strong attachment to its human owners. In consequence, it appears plausible that participants may have felt more familiarity and closeness toward Spot compared with Atlas. Thus, our present findings can be reconciled with previous work on anthropomorphism and mind perception (e.g. Gray and Wegner, 2012) if we allow that certain types of zoomorphism might tap even more strongly into the underlying mechanisms of the uncanny valley.

Our results have implications for the design and deployment of social robots, as both developers and users need to decide how best to balance advantages of robots that can engage us in interaction through socio-emotional bonding (e.g. Jones et al., 2015), and potentially negative effects of user discomfort arising from empathy for an entity that we intuitively start to treat like another person. We already see versions of this playing out in the deployment, marketing, and adoption of digital assistants for the home, and it is likely that similar opportunities/challenges will accompany the introduction of more sophisticated social robots in domestic settings. However, while certain types of proximity or closeness may breed discomfort, the lower levels of discomfort toward the Roomba suggest that this effect may disappear over time, for example, due to repeated interactions (Złotowski et al., 2018), or more long-term changes in familiarity with increasingly intelligent technological artifacts. Thus, while we do not have the data to show what levels of discomfort such a cleaning robot would have induced 20 years ago, it stands to reason that it would have been substantially higher. Furthermore, this consideration emphasizes our argument that more research is needed to examine the impact of current technological developments and their impact on everyday life.

Attitudes toward AI

Our findings demonstrate an intriguing association between the robots' morphology and functions and user attitudes. The more advanced or capable a robot appears to be, the greater the level of user hopes and fears regarding the technology. The Roomba, for instance, does not elicit the same level of hopes and fears as does the Spot robot from Boston Dynamics. This difference may be due to the fact that the Roomba, in both its design and its functionality, remains closer to a household tool than it is a socially situated subject who counts as something Other. As already mentioned, it is now also a much more familiar and unobtrusive item than the novel and partially unpredictable capabilities demonstrated by the other two robots. Unfortunately, based on the limited scope of our study, there is still insufficient data to evaluate whether these differences are due to

the nature of the robot, are an artifact of the video representation of the robot, or result from a combination of the two. Responding to this uncertainty necessitates careful consideration of the one significant limitation encountered by the investigation.

Conclusions

Empirical research in HRI is typically divided into one of two types: Simulated robot studies, including picture-viewing paradigms (e.g. Von der Pütten and Krämer, 2012) and real-world robot studies. The latter is predicated upon human subjects interacting with actual, physically embodied robots in a laboratory or controlled real-world setting. The former employs various mediated representations of robots—written descriptions, photographs, non-player characters within the virtual space of computer games, or videos, as was the case with our investigation. There are recognized advantages and disadvantages to simulated robot studies (cf. Broadbent, 2017). Consequently, it might be prudent to repeat the present investigation using actual robots and having test subjects either witnessing the robot abuse in person, or even engaging in seemingly abusive practices. Whereas the former study would involve substantial costs and challenges to creating comparable conditions of harm, the latter design would additionally have to carefully address ethical concerns in case participants might be required to act against their conscience. Thus, the next step in this program of research could be to do similar investigations of robot abuse in an augmented reality (AR) setting wherein realistic virtual robots are harmed as part of a natural and ecologically valid setting, such as in the homes of participants.

Despite this need for further study, the present study addresses a noteworthy gap between the two types of traditional research designs in this field: By exposing participants to impressive examples of robots and robot abuse on the cutting edge of current robot design, our study could examine the effects of harmful interaction with surprisingly capable new robots. Compared with previous experimental work using toy robots (e.g. a Pleo; Rosenthal-von der Pütten et al., 2013), this implies that participants saw and evaluated materials featuring a testing procedure in which the context of the actions of both the robots and the human personnel were natural in the sense that the harm done to the robots served a clearly identifiable and not artificially generated purpose. Thus, the present study did not require a cover story or other arbitrary information to explain to participants why someone might have created such videos or how they might be expected to respond to be "good participants" (demand effects). Notably, our selection of robots furthermore avoided, or at least reduced, certain potential confounds associated with the use of toy robots: The robots were similar in the sense that they were all clearly mechanical in nature. Spot and Atlas, that is, the two robots of immediate interest for our research question concerning the harm-made mind, were produced by the same manufacturer, made of similar materials, and underwent comparable types of testing as featured in the videos shown to our participants. Likewise, compared with previous work (e.g. Rosenthal-von der Pütten et al., 2013) the Roomba was certainly not a squishy toy designed to elicit emotions from customers. It therefore appears unlikely that our present findings might simply be explained by basic differences in apparent vulnerability, softness, or cuteness. And while control over these factors could not be perfect, overall

comparability between these robots was likely higher than in other studies featuring the robots available at a given laboratory. Thus, consistent with notions of relational ethics (Gunkel, 2018), the present work provides novel and systematic evidence on human perception of highly advanced robots that we believe takes an important step beyond the often rather anecdote-driven discourse concerning the perceived moral standing of robots as social entities. Likewise, while perfect control over all potentially relevant factors remained unattainable, our approach enabled us to study reasonably comparable responses to highly advanced robots that are not (yet) generally available to social robotics research. Our results further provide a new perspective on previous anecdotal evidence, suggesting that even "hardened" robots such as Atlas and Spot may be able to push our own social buttons in a way that is not yet sufficiently understood.

Finally, we hope that our results concerning the perception and attitudes toward robots and AI will provide a starting point for further research investigating the impact of mass media and online depictions of current advances in robotics and artificial intelligence. Here, we were surprised to what extent short videos on YouTube appeared to shape general attitudes toward future AI. Certainly, even though the behaviors showcased for Atlas and Spot are impressive compared with more familiar robots, they are still a far cry from the capabilities ascribed to robots in science fiction. In this sense, our results could be interpreted as good news, because participants still appeared to be open to attitudinal change. Watching these videos could nevertheless have triggered expectations associated with fictional robots, either via explicit reflection on future AI in view of new evidence, or via processes of implicit priming. Overall, our study thus highlights how we can be responsive to harm done to robots—while simultaneously feeling discomforted and concerned. This apparent paradox (Vanman and Kappas, 2019) further highlights the need for methodological advancement in this field. We should make good use of this time, where robots as social interaction partners are still an emerging new "other," so that we might not again (cf. e.g. Colella et al., 2017) trail too far behind the times when insights from theory need to be applied in the real world. A combination of qualitative and quantitative approaches on openly available online videos might provide one such new piece of the puzzle.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a grant Sonata (2016/23/D/HS6/02954) from the Polish National Science Centre (Narodowe Centrum Nauki), and by the German Research Foundation (DFG) through the project "PALMS" (no. 402779631), which is a subproject of the DFG Priority Program "The active self" (SPP 2134; no. 360292909).

ORCID iD

Dennis Küster https://orcid.org/0000-0001-8992-5648

Notes

1. We fell slightly short of our recruitment target for 90% power, with the obtained *N* of 160 participants (40 per condition) being equivalent to 88% power. However, we considered this difference as within the margin of error given that our power analysis reflected conventional parameters for medium-sized effects, where no exact empirical data could be obtained from prior work.

2. Respectively for Spot and Atlas, the following Youtube videos were used: https://www.youtube.com/watch?v=fUyU3lKzoio and https://www.youtube.com/watch?v=fRj34o4hN4I, and https://www.youtube.com/watch?v=aFuA50H9uek&t=3s and https://www.youtube.com/watch?v=rVlhMGQgDkY as versions with harm depiction.
3. https://www.youtube.com/watch?v=9f6whsU9m3M
4. Participant gender was included as a covariate. However, although the effect of gender was marginally significant ($p = .078$), it did not alter the pattern of results and we thus excluded it from further analyses, unless stated otherwise.
5. The main effect of robot type was significant for experience ($M_{Spot} = 2.06$, $SD = 1.30$ vs $M_{Atlas} = 1.57$, $SD = .88$), $F(1, 125) = 6.33$, $p = .013$, $\eta_p^2 = .05$; agency ($M_{Spot} = 3.76$, $SD = 1.09$ vs $M_{Atlas} = 3.09$, $SD = 1.09$), $F(1, 125) = 12.19$, $p = .001$, $\eta_p^2 = .09$; and consciousness ($M_{Spot} = 3.45$, $SD = 1.78$ vs $M_{Atlas} = 2.72$, $SD = 1.58$), $F(1, 125) = 6.66$, $p = .011$, $\eta_p^2 = .05$, but not for pain ($M_{Spot} = 1.75$, $SD = 1.51$ vs $M_{Atlas} = 1.50$, $SD = 1.02$), $F(1, 125) = 1.22$, $p = .272$, $\eta_p^2 = .01$.
6. Gender as a significant covariate ($p = .001$) did not alter the pattern of results.
7. Participant gender was a marginally significant covariate, $p = .061$.

References

Banks J (2020) Theory of mind in social robots: replication of five established human tests. *International Journal of Social Robotics* 12: 403–414.

Bartneck C and Hu J (2008) Exploring the abuse of robots. *Interaction Studies* 9(3): 415–433.

Bartneck C, Brahnam S, De Angeli A, et al. (2008) Misuse and abuse of interactive technologies. *Interaction Studies* 9(3): 397–401.

Bartneck C, Reichenbach J and Carpenter J (2006) Use of praise and punishment in human-robot collaborative teams. In: *ROMAN 2006—the 15th IEEE international symposium on robot and human interactive communication*, Hatfield, 6–8 September, pp. 177–182. New York: IEEE.

Bartneck C, Rosalia C, Menges R, et al. (2005) Robot abuse—a limitation of the media equation. In: *Proceedings of the INTERACT '05 workshop on agent abuse* (eds A De Angeli, S Brahnam and P Wallis), Rome, 12 September, pp. 54–57. Available at: http://hdl.handle.net/10092/16925

Brahnam S and De Angeli A (2008) Special issue on the abuse and misuse of social agents. *Interacting with Computers* 20(3): 287–291.

Breazeal CL (2002) *Designing Sociable Robots*. Cambridge, MA: MIT Press.

Broadbent E (2017) Interactions with robots: the truths we reveal about ourselves. *Annual Review of Psychology* 68(1): 627–652.

Carpinella CM, Wyman AB, Perez MA, et al. (2017) The robotic social attributes scale (RoSAS): development and validation. In: *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*, Vienna, 6–9 March, pp. 254–262. New York: ACM.

Ceh S and Vanman EJ (2018) The robots are coming! The robots are coming! Fear and empathy for human-like entities. *PsyArXiv*. Epub ahead of print 4 June. DOI: 10.17605/OSF.IO/4CR2U.

Coeckelbergh M (2012) *Growing Moral Relations: Critique of Moral Status Ascription*. New York: Palgrave Macmillan.

Colella A, Hebl M and King E (2017) One hundred years of discrimination research in the Journal of Applied Psychology: a sobering synopsis. *Journal of Applied Psychology* 102(3): 500–513.

Darling K (2015) Robot ethics is about humans. Available at: http://videos.theconference.se/robots-and-humans (accessed 21 January 2020).

De Angeli A, Brahnam S and Wallis P (2005) ABUSE: the dark side of human-computer interaction. In: *Adjunct proceedings of the INTERACT '05 IFIP Tc13 international conference on human-computer interaction*, Rome, 12–16 September, pp. 91–92. Available at: http://www.agentabuse.org

De Angeli A, Brahnam S, Wallis P, et al. (2006) Misuse and abuse of interactive technologies. In: *Extended abstracts on human factors in computing systems (CHI'06)*, Montreal, QC, Canada, April, pp. 1647–1650. New York: ACM.

Derrida J (2005) *Paper Machine* (trans. R Bowlby). Stanford, CA: Stanford University Press.

Edwards AC, Edwards PR, Spence D, et al. (2016) Initial interaction expectations with robots: testing the human-to-human interaction script. *Communication Studies* 67(2): 227–238.

Epley N, Waytz A and Cacioppo JT (2007) On seeing human: a three-factor theory of anthropomorphism. *Psychological Review* 114(4): 864–886.

Fast E and Horvitz E (2017) Long-term trends in the public perception of artificial intelligence. In: *Proceedings of the thirty-first AAAI conference on artificial intelligence*, San Francisco, CA, 4–9 February, pp. 963–969. Palo Alto, CA: AAAI Press.

Faul F, Erdfelder E, Lang AG, et al. (2007) G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39(2): 175–191.

Floridi L (2013) *The Ethics of Information*. Oxford: Oxford University Press.

Gray HM, Gray K and Wegner DM (2007) Dimensions of mind perception. *Science* 315(5812): 619.

Gray K and Wegner DM (2009) Moral typecasting: divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology* 96(3): 505–520.

Gray K and Wegner DM (2012) Feeling robots and human zombies: mind perception and the uncanny valley. *Cognition* 125(1): 125–130.

Gray K, Knickman TA and Wegner DM (2011) More dead than dead: perceptions of persons in the persistent vegetative state. *Cognition* 121(2): 275–280.

Gray K, Young L and Waytz A (2012) Mind perception is the essence of morality. *Psychological Inquiry* 23(2): 101–124.

Gunkel DJ (2018) *Robot Rights*. Cambridge, MA: MIT Press.

Heider F and Simmel M (1944) An experimental study of apparent behavior. *The American Journal of Psychology* 57(2): 243–259.

Jones A, Küster D, Basedow CA, et al. (2015) Empathic robotic tutors for personalised learning: a multidisciplinary approach. In: Tapus A, André E, Martin JC, et al (eds) *Social Robotics. International conference on social robotics (ICSR), Lecture Notes in Computer Science, vol. 9388*. Cham: Springer, pp. 229–285.

Jones KS and Schmidlin EA (2011) Human-robot interaction: toward usable personal service robots. *Reviews of Human Factors and Ergonomics* 7(1): 100–148.

Jones T, Lawson S and Mills D (2008) Interaction with a zoomorphic robot that exhibits canid mechanisms of behaviour. In: *2008 IEEE international conference on robotics and automation*, Pasadena, CA, 19–23 May, pp. 2128–2133. New York: IEEE.

Kahn PH Jr, Kanda T, Ishiguro H, et al. (2012) Do people hold a humanoid robot morally accountable for the harm it causes? In: *Proceedings of the 7th annual ACM/IEEE international conference on human–robot interaction (HRI'12)*, Boston, MA, March, pp. 33–40. New York: ACM.

Konok V, Korcsok B, Miklósi Á, et al. (2018) Should we love robots? – The most liked qualities of companion dogs and how they can be implemented in social robots. *Computers in Human Behavior* 80: 132–142.

Löffler D, Dörrenbächer J and Hassenzahl M (2020) The uncanny valley effect in zoomorphic robots: the U-shaped relation between animal likeness and likeability. In: *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*, pp. 261–270. Available at: https://doi.org/10.1145/3319502.3374788

Loughnan S, Pina A, Vasquez EA, et al. (2013) Sexual objectification increases rape victim blame and decreases perceived suffering. *Psychology of Women Quarterly* 37(4): 455–461.

MacDorman KF, Green RD, Ho C-C, et al. (2009) Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior* 25(3): 695–710.

Miklósi Á and Gácsi M (2012) On the utilization of social animals as a model for social robotics. *Frontiers in Psychology* 3: 75.

Moon Y (2000) Intimate exchanges: using computers to elicit self-disclosure from consumers. *Journal of Consumer Research* 26(4): 324–340.

Moore RK (2012) A Bayesian explanation of the "Uncanny Valley" effect and related psychological phenomena. *Scientific Reports* 2(1): 864.

Mori M, MacDorman KF and Kageki N (2012) The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine* 19(2): 98–100.

Nass C and Moon Y (2000) Machines and mindlessness: social responses to computers. *Journal of Social Issues* 56(4): 81–103.

Nass C, Moon Y and Carney P (1999) Are respondents polite to computers? Social desirability and direct responses to computers. *Journal of Applied Social Psychology* 29(5): 1093–1110.

Nass C, Moon Y, Morkes J, et al. (1997) Computers are social actors: a review of current research. In: Friedman B (ed.) *Moral and Ethical Issues in Human-Computer Interaction*. Stanford, CA: CSLI Press, pp. 137–162.

Parke P (2015) Is it cruel to kick a robot dog? Available at https://edition.cnn.com/2015/02/13/tech/spot-robot-dog-google/index.html (accessed 21 January 2020).

Reeves B and Nass C (1996) *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York: Cambridge University Press.

Rosenthal-von der Pütten AM, Krämer NC, Hoffmann L, et al. (2013) An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics* 5(1): 17–34.

Sandry E (2015) *Robots and Communication*. New York: Palgrave Macmillan.

Suzuki Y, Galli L, Ikeda A, et al. (2015) Measuring empathy for human and robot hand pain using electroencephalography. *Scientific Reports* 5: 15924.

Swiderska A and Küster D (2018) Avatars in pain: visible harm enhances mind perception in humans and robots. *Perception* 47(12): 1139–1152.

Vanman EJ and Kappas A (2019) "Danger, Will Robinson!" The challenges of social robots for intergroup relations. *Social and Personality Psychology Compass* 13(8): 1–13.

Von der Pütten AM and Krämer NC (2012) A survey on robot appearances. In: *Proceedings of the seventh annual ACM/IEEE international conference on human-robot interaction (HRI '12)*, Boston, MA, March, pp. 267–268. New York: ACM.

Von der Pütten AM, Krämer NC, Gratch J, et al. (2010) "It doesn't matter what you are!": explaining social effects of agents and avatars. *Computers in Human Behavior* 26(6): 1641–1650.

Ward AF, Olsen AS and Wegner DM (2013) The harm-made mind: observing victimization augments attribution of minds to vegetative patients, robots, and the dead. *Psychological Science* 24(8): 1437–1445.

Waytz A, Gray K, Epley N, et al. (2010) Causes and consequences of mind perception. *Trends in Cognitive Sciences* 14(8): 383–388.

Złotowski J, Proudfoot D, Yogeeswaran K, et al. (2015) Anthropomorphism: opportunities and challenges in human–robot interaction. *International Journal of Social Robotics* 7(3): 347–360.

Złotowski JA, Sumioka H, Nishio S, et al. (2018) Persistence of the uncanny valley. In: Ishiguro H and Dalla Libera F (eds) *Geminoid Studies*. Singapore: Springer, pp. 163–187.

Author biographies

Dennis Küster is a post-doctoral researcher at the Department of Computer Science, University of Bremen, Germany. His interdisciplinary research focuses on emotions and social cognition in Human-Computer-Interaction, and aims to better understand how we attribute mind and morality to (social) robots.

Aleksandra Swiderska, PhD, works as an assistant professor at the Department of Psychology, University of Warsaw, Poland. She is interested in social cognition and studies the processes of anthropomorphism (perceiving objects as people) and dehumanization (perceiving people as objects).

David Gunkel is distinguished teaching professor at Northern Illinois University. He is the author of twelve books on digital media technology and ethics, including *The Machine Question* (MIT Press 2012) and *Robot Rights* (MIT Press 2018).