Research Article

Christoph Bartneck* and Merel Keijsers

The morality of abusing a robot

https://doi.org/10.1515/pjbr-2020-0017
received December 16, 2019; accepted May 9, 2020

Abstract: It is not uncommon for humans to exhibit abusive behaviour towards robots. This study compares how abusive behaviour towards a human is perceived differently in comparison with identical behaviour towards a robot. We showed participants 16 video clips of unparalleled quality that depicted different levels of violence and abuse. For each video, we asked participants to rate the moral acceptability of the action, the violence depicted, the intention to harm, and how abusive the action was. The results indicate no significant difference in the perceived morality of the actions shown in the videos across the two victim agents. When the agents started to fight back, their reactive aggressive behaviour was rated differently. Humans fighting back were seen as less immoral compared with robots fighting back. A mediation analysis showed that this was predominately due to participants perceiving the robot's response as more abusive than the human's response.

Keywords: abuse, robots, human, morality, perception

1 Introduction

The interactions between humans and robots are not always positive. Some humans verbally and physically abuse robots. Hitchbot, for example, was completely destroyed in Philadelphia, Pennsylvania, when it was trying to catch a lift [1]. Robovie, which was operating in kindergartens, schools, shopping malls, and train stations, had to deal with a variety of abusive behaviours, including people obstructing its way and kicking it [2]. The K5 Knightscope robot was assaulted by a drunk man in a parking lot in Mountain View, California [3]. How to respond to this abusive behaviour is a difficult question, and researchers are trying to run controlled experiments to better understand people's motivation for abusing robots and what the best response strategies would be.

One approach to understanding robot bullying has been to experimentally study what makes people aggressive towards robots [4–6]. However, experiments designed so, i.e. participants have to verbally or physically abuse a robotic agent, may have a hard time passing the ethics board. In addition, people are unlikely to bully robots during a controlled experiment, as they tend to be self-aware and wanting to make a good impression [7,8]. People have been coerced to physically harm a robot [9], but in this experiment participants were explicitly instructed to destroy the robots and the robots in question were very simple and cheap. Therefore, the behaviour that was measured may have been obedience rather than robot abuse. It is uncertain whether these results would generalise to robot bullying, or even obedience behaviour with a more advanced and anthropomorphic robot. In short, studies targeting robot abuse are complicated to design, and researchers often have to adopt a proxy for abusive behaviour.

An alternative approach is to expose participants to recordings of robot abuse and measure their responses. Previous studies have measured to what extent participants perceived the shown behaviour to be abusive [10–12], empathy with the robot [13,14], and acceptability of the abuse [15]. No studies to date, to our knowledge, have directly compared the morality of the abuse of robotic versus human agents. However, such a comparison is needed to place measures into perspective. If a study only observed abusive behaviour towards a robot and concluded with a "three-point-oh-seven" on a five-point empathy scale, then this benchmark is not meaningful by itself. The current study therefore compares the moral acceptability of human and robot abuse directly as well as the moral acceptability of reactive aggression.

1.1 Literature

Humans respond to robots as if they are to some extent sentient and humanlike. This goes as far as parallels in how the brain responds to human–human and human–

* **Corresponding author: Christoph Bartneck,** HIT Lab NZ, University of Canterbury, Private Bag 4800, 8140 Christchurch, New Zealand, e-mail: christoph.bartneck@canterbury.ac.nz
Merel Keijsers: HIT Lab NZ, University of Canterbury, Private Bag 4800, 8140 Christchurch, New Zealand, e-mail: merel.keijsers@pg.canterbury.ac.nz

robot interaction (HRI) [13,16,17], but it also shows in people's tendency to interact with robots in a social manner [18]. Humans talk to robots [19] as if they understand what is being said, punish them for being a bad teammate [20] but also feel sorry for them when they are being punished [21] and even try to prevent them from getting hurt [21,22].

Not all social behaviours are positive. Social robots have been the targets of verbal and physical abuse in the past [1,3,23,24]. What is even more interesting is that robot-directed aggression has been shown to be remarkably persistent [see, e.g. 6,23–25], and researchers have been struggling to come up with adequate robot responses to effectively deter further abusive behaviour. Robots are in many ways the ideal target for abuse, as they are in a clear subordinate position, are not expected to retort in kind, and cannot feel any pain, which absolves the aggressor from any moral consequence [26].

This is not to say that robot bullying should be tolerated. From an ethical perspective, some behaviour can be deemed immoral even if it is performed on an entity that is incapable of any suffering, like a robot [27]. Since robots are recognised by humans as social actors, abusing them might encourage treating humans in a similar way [28]. More generally speaking, the assertion "I can do whatever I desire with a robot" rests upon the idea that all and any actions are acceptable as long as no one gets harmed [29], which even in the most libertarian societies is not a commonly shared attitude [28]. And from a pragmatic point of view, robot abuse can result in considerable damage to the robot and hazardous situations for the robot, the abuser, any bystanders, and future users [26].

Researchers are facing several methodological problems when trying to investigate abusive behaviour towards robots. The biggest practical problem is that the physical abuse could damage or even destroy the robots. Conducting experiments that involve physically abusing robots could therefore be prohibitively expensive, unless a very cheap robot is being used [9,30]. These tend to be very simple in terms of behaviour and appearance, which would potentially bias the results and prevent generalising any findings to more advanced and anthropomorphic robots.

To overcome these problems, researchers have resorted to less destructive forms of abuse with more sophisticated robots. For instance, robot bullying has been operationalised as reducing a robot's electrical power supply [4,19] or using abusive language [5]. Using the framework of Game Theory, robots have also been withheld points or money [31].

These milder forms of abuse could also be employed in studies where abuse of robots is compared with abuse of humans. Withholding money from participants is a methodology that could still pass an ethics board. Kicking and hitting a participant are not. Any studies comparing human-robot abuse and human-human abuse would therefore be limited to mild forms of abusive behaviour. Aggression towards humans can be measured through small transgressions of social norms; rude behaviour that won't cause any physical or severe psychological harm but is enough to slightly sting. Robots however are not sentient and most people are rationally well aware of this. Previous research has suggested that this does not prevent people from automatically being polite [32], but one could still argue that a participant omitting any polite conversation or withholding any reward (monetary or otherwise) from a robot does so because they reasoned that the robot could not care less about whether a command ends with "please" or if it is awarded any payment. Thus, rather than robot abuse being measured, it could be the participant's desire to come across as a rational human being.

Alternatively, more extreme abusive behaviours can be studied through registering participant's responses to recordings of abuse. Common examples of this method are vignette-based approaches, where participants read about abusive behaviour and then express their moral sentiment towards the actions described [e.g. 33], video-based approaches, where participants are shown a film clip of robot abuse [see, for instance, 14] or indicating behavioural intentions after interaction with a robot [15].

The discussion about whether representations of HRI, such as videos and texts, could be used as valid stimuli in HRI studies is ongoing. Robert Sparrow argued that such representations have sufficient moral value to serve as a test for the humans' virtue [27]. Previous studies have shown that virtual representations of robots elicit more social behaviour (e.g. mimicking expressions, feelings of empathy, polite behaviour, and physiological responses) than audiotapes or text [13,21], indicating that virtual robots, too, are recognised as social agents. Li [34] conducted a meta-analysis on papers that studied the influence of agent embodiment on users' perception of the agent and concluded that embodied robots elicit stronger behavioural and attitudinal responses than virtual agents. Several studies that had found no difference in behavioural and attitudinal responses for virtual agents and physical robots were missing in this analysis, however [e.g. 35,36]. More recent studies, as well, found that the perception of and response to virtual agents is identical to embodied robots [37,38]. More specifically, Thellman et al. [38] found that social presence (i.e. whether the robot is perceived as a social actor that manifests humanness [39]) rather than physical presence predicts social influence of a robot. In their experiment, social presence was not influenced by the physical embodiment of the robot. At the same time, Keijsers et al. [4] found that robot embodiment had an effect on people's willingness to administer punishments: embodied robots got less severe punishment than their virtual replica. The discussion is, in other words, still ongoing. While studies seem to confirm that virtual agents can definitely elicit social responses, the question remains whether these are as intense as they would have been with an embodied robot. That being said, there is little doubt that virtual representations of robots can elicit emotional responses.

This was demonstrated as well by the public response to a video of a man kicking a robot dog. In February 2015, Boston Dynamics published a video of its quadruped robot "Spot". Employees kicked the robot in order to demonstrate the robot's capacity to regain its balance.[1] The video went viral and sparked discussions about the morality of the demonstrated behaviour [40], with many commenters perceiving the kicks to be abusive [30]. In other videos, Boston Dynamics employees used a hockey stick to remove a box from the grip of Atlas, a humanoid robot.[2] The intention was to demonstrate Atlas' capacity to dynamically track and grip a box. Many viewers of the video considered it as teasing behaviour that is abusive. Boston Dynamics has since included a disclaimer to their robot videos to assure viewers that the behaviour "does not irritate or harm the robot" [41].

It seemed therefore feasible to study how people responded to robot abuse by collecting their responses to video recordings of more extreme cases of robot abuse than would be possible to set up in a lab experiment. Comparing or even benchmarking these responses to how people react to humans being exposed to the same abusive behaviour remained, however, more problematic. Up until now, no stimuli were available that would convincingly show the exact same abusive behaviour towards a robot and towards a human. Comparison studies were therefore often constrained to text stimuli.

1.2 Current study

On 15 June 2019, Corridor Digital published a video, in which an Atlas robot was shown as it performed a number of tasks while a human engineer deliberately attempted to sabotage them. These sabotaging behaviours got gradually more aggressive, until the robot turned and attacked the "bullying" human.[3] The robot in this video was computer-generated imagery (CGI) rendered; its motions had been captured through a human in a tracksuit. As a result, there were two versions of a video with identical abusive behaviours: one video where the victim was a human and one where it was a robot. This unique footage allowed us to compare the perceived morality of the exact abusive behaviour when carried out towards a human versus a robot. See Figure 1 for a side-by-side comparison of the same frame for the human and the robotic agent.

In the current experiment, participants watched 14 instances of abusive behaviour towards either the robot or the human agent and indicated how morally (un) acceptable they perceived these behaviours to be. After the 14 videos that showed aggression towards the agent, two additional video clips were shown where the agent started fighting back (i.e. reactive aggression). Thus, the moral acceptability of reactive aggression to the group that just abused the agent was assessed.

1.2.1 Methodological considerations

The original videos showed how a group of human bullies abuse either a single human or a single robot. A straightforward experimental set-up would be to show one group of participants the videos in which the human gets abused and the other group the videos in which the robots gets abused.

Towards the end of the original video, however, the victim starts to fight back. Initially, it grabs the hockey stick, yanks it out of the hands of the bully, and throws it to the ground. Later, the victim hits and kicks the human bullies. If this storyline is retained in the experiment, it becomes possible to talk about bullies, victims, and fighting back. It would be interesting to measure whether the violent acts of the victim fighting back would be considered as less morally problematic than the acts of unprovoked aggression by the bullies before. The fighting back could be considered as a form of self-

[1] https://youtu.be/M8YjvHYbZ9w

[2] https://youtu.be/rVlhMGQgDkY

[3] https://youtu.be/dKjCWfuvYxQ

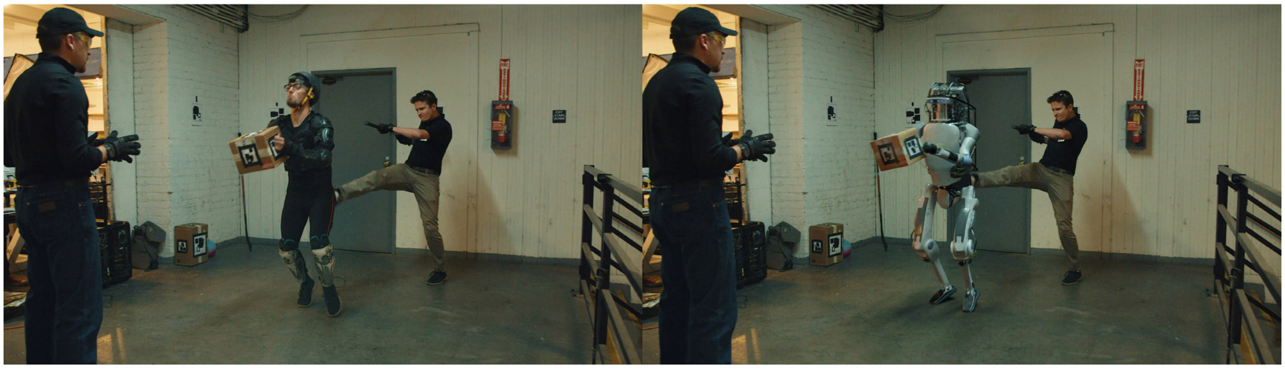


Figure 1: Kicking a human and a robot in the back.

defence. We define acts of violence that are a response to previous bullying as *reactive aggression*.

We wanted to go one step further than the straightforward experimental set-up described above and take the reactive aggression scenes into consideration as well. For a systematic experiment, we would need to counterbalance the sequence in which these two types of videos are presented. We had 14 videos in which a group of human bullies abuse a single human/robot, and we had 2 videos in which the single human/robot abuses a group of humans. One group would see the 14 videos first and the other group would see the 2 videos first.

However, there are several methodological problems with such a manipulation. The most obvious one would be the imbalance in the number of videos. Any reactive aggression after being provoked on 14 previous occasions would most likely be perceived differently than an aggressive response to only 2 previous abuses. There are other confounding factors, like the group size. The 14 videos show a group of up to three humans bullying an individual, while the 2 videos show an individual fight against a group of humans. Another confounding factor is the intensity of the abuse. The 14 videos include scenes where the bullies fire a gun which would almost certainly be seen as more abusive than tossing away a hockey stick or kicking a person. Moreover, a complete factorial design would be impossible as the human(s) in the videos can be seen abusing a robot and other human(s), but the robot can never be seen abusing another robot and there is never more than one robot in the scenes. Several potential research questions, such as comparing unprovoked aggression to reactive aggression for the same agent, were therefore unfortunately difficult to study.

So we had to conceptualise the experiment in a different way. Besides the original comparison between acceptability of abuse depending on whether the victim is a human or a robot, we also studied moral acceptability of reactive aggression to the group that just abused the agent. This was done by showing the two reactive aggression videos after the 14 abusive videos and measuring moral acceptability and perceived abuse, violence, and intent to harm. While we did not systematically manipulate the presentation order, this design still allows some specific research questions to be answered.

1.2.2 Research questions

Given these considerations, we were interested in the following research questions:

1. Is abusing a robotic agent seen as more morally acceptable than abusing a human?
2. Is reactive aggression more acceptable when it comes from the human agent than from the robotic one?
3. If abusive behaviour in one agent condition is seen as more acceptable than the other, is this difference in acceptability due to a different perception of how abusive the behaviour was perceived to be?
4. Do the different abusive behaviours cluster based on their perceived violence and intention to hurt?

Based on the media equation theory [32,42] as well as empirical evidence that viewing similar (although not identical) abusive behaviours towards robots and humans elicits similar neurological responses [14], it was expected that the abuse of the robotic agent would not be considered more acceptable than the abuse of the human agent. Since it was expected that abusive behaviour to both agents would be seen as equally unacceptable, it was furthermore expected that there would be no difference in how acceptable reactive aggression from the agent was seen.

2 Method

2.1 Participants and design

The experiment followed a single-factor (agent: human or robot) between-subject design. Participants watched (in randomised order) 14 videos in which an agent was exposed to various types of abuse. After each video, they rated the behaviour shown in the video clip for moral acceptability. These 14 videos were then followed by two videos in which the agent fought back. These two videos as well were each rated for moral acceptability. Finally, the participants filled out questionnaires on mind attribution to the agent, individual tendency to anthropomorphise, and affinity with technology.

The dependent variables (DVs) were perceived acceptability of the videos where the agent was abused, perceived acceptability of the videos where the agent fought back, and mind attribution to the agent. Individual tendencies to anthropomorphise and affinity with technology were used as randomisation checks. This study was approved by the University of Canterbury Ethics board under the reference HEC 2019/30/LR-PS.

In all, 166 participants were recruited from Amazon Mechanical Turk (MTurk). Previous studies have indicated that data collected via MTurk are of equal quality to on-campus recruitment or participant data from forums [43,44], with internal motivation rather than monetary reward being the main motive for participating [45]. Participants received US$1 for their participation, which is in line with the MTurk reimbursement custom. The survey took approximately 10 min to complete.

All participants were native English speakers and lived in the USA, UK, Canada, Ireland, Australia, or New Zealand. All participants were Amazon Mechanical Turk Master Workers. These workers are being monitored by Amazon for their performance over time. Amazon explains that "Workers who have demonstrated excellence across a wide range of tasks are awarded the Masters Qualification. Masters must continue to pass our statistical monitoring to retain their qualification".

Of the 166 participants, 5 reported being familiar with the video material but did not think the video was unrealistic. Thirty participants had not seen the clip but thought the material was unrealistic. Finally, four reported having seen the video before and thinking the material was unrealistic. All 39 participants were removed from the data set, resulting in a final data set of 127 participants. In all, 51.18% (n = 65) were male; the mean age was 42.57 years (SD = 11.20; range = 25–72). Fifty-nine participants saw the human agent videos, while the other 68 were in the robotic agent condition.

2.2 Measurements

2.2.1 Moral acceptability of abuse

We measured the moral acceptability of the aggressive behaviour shown in the video clip through a single item assessed after each video. The item stated *How (un)acceptable would you say the behaviour shown in the video is?* Participants could indicate their answer on a seven-point scale which consisted of the following answer options (see also Figure 2):

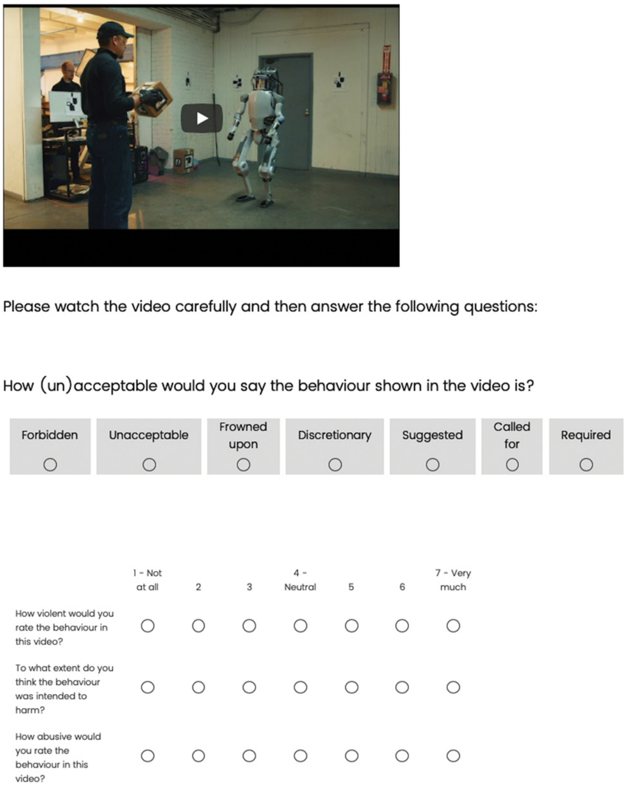
Forbidden
Unacceptable
Frowned upon
Discretionary
Suggested
Called for
Required

The terminology for the response options was taken from work on the dimensions of normative demand by Malle et al. [46]. Malle et al. validated a scale that held 13 points and ranged from prescriptions to prohibitions. To keep the scale readable for participants and avoid any formatting issues that could occur when an extensive scale would be displayed on a smaller screen (e.g. of a tablet or laptop), we reduced the number of items to seven by omitting every other point on the original scale. Previous analyses have indicated that from 5 to 7 points on (depending on the covariance between the items) adding extra points to a scale does not alter the reliability of a scale [47].

Since acceptability was measured with a single item only, its construct validity was assessed by correlating it with perceived violence, abusiveness, and intention to harm. Based on Cohen [48], Pearson correlation coefficients of 0.6 or higher (i.e. a large effect size) were expected.

2.2.2 Violence, abusiveness, and intention to harm

Participants were furthermore asked to rate each video on three additional scales: how violent they thought the behaviour was, to what extent the behaviour had been intended to harm, and how abusive the behaviour was. Each item was answered on a 7-point scale ranging from *Not at all* to *Very much*. See Figure 2 for a screenshot of one of the videos plus the four questions.



Please watch the video carefully and then answer the following questions:

How (un)acceptable would you say the behaviour shown in the video is?

Forbidden	Unacceptable	Frowned upon	Discretionary	Suggested	Called for	Required
○	○	○	○	○	○	○

	1 - Not at all	2	3	4 - Neutral	5	6	7 - Very much
How violent would you rate the behaviour in this video?	○	○	○	○	○	○	○
To what extent do you think the behaviour was intended to harm?	○	○	○	○	○	○	○
How abusive would you rate the behaviour in this video?	○	○	○	○	○	○	○

Figure 2: Screenshot of one of the videos plus the questionnaire.

2.2.3 Individual differences in anthropomorphism

After having seen and rated all 16 videos, participants completed the individual differences in anthropomorphism questionnaire [49], which measured their personal tendency to attribute different aspects of sentience and various emotions to a wide range of non-human entities (e.g. natural phenomena and animals). The original scale includes items that refer to mechanical entities as well (e.g. robots, cars, and computers). For the current experiment, these were omitted since the experimental manipulation could bias responses on those items. The resulting questionnaire held 10 questions like "To what extent does a fish have free will" and "To what extent does the environment have emotions", which were rated on a 5-point Likert scale ranging from *Not at all* to *Very much*. Individual differences in anthropomorphism were used as a randomisation check between the conditions.

2.2.4 Control questions

Finally, two control questions at the very end of the survey were included: (1) Have you seen this particular video before? and (2) How authentic were the movie clips (on a seven-point scale, ranging from *Obviously not realistic* (*animated*) to *Clearly realistic*)? Participants who responded with *Definitely yes* or *Probably yes* to the first question, or the lower end (i.e. 1, 2, or 3) of the realism scale, were excluded from the analyses.

2.3 Video material

On 15 June 2019, Corridor Digital published a video in which the Boston Dynamics' Atlas robot was shown executing various attempts of picking up and carrying around a cardboard box under supervision of a human engineer. In the video, another human performed a variety of abusive behaviours towards the robot. These started with what Boston Dynamics had shown in its original videos, such as kicking the agent or using a hockey stick to interfere with the grabbing of the box. Over time, the behaviours got increasingly abusive and peaked as the human bully shot the robot with a gun. Eventually, the robot started fighting back and the roles got reversed. The robot forced its human bullies to carry boxes for it by holding them at gunpoint.

This special effects video was extremely well done and fooled nearly everyone to believe that an actual Atlas robot was used, while it was in reality a computer-generated model. Corridor Digital used motion tracking of a human actor to capture the behaviour and mapped a digital Atlas robot onto the movements to create the animation. Upon request, they kindly shared both the motion-capturing footage, showing the human actor and the special effects video with the Atlas robot (see Figure 1 for a side-by-side comparison between the same frame from the unedited video and the complete special effects video).

Each of the two versions was cut into 16 video clips; 14 of those depicted abusive behaviour towards the agent and 2 showed the agent responding with aggression to the human engineers. The 14 abusive videos showed a wide range of aggressive behaviours. Three scenes did not include any physical abuse but instead showed verbal abuse, such as "You are completely useless!" The other 11 video clips showed physical abuse or taunting.

The resolution of the videos was reduced from 3,840 × 2,160 pixels to 1,280 × 720 pixels to ensure fast playback on mobile devices. Video playback speed was tested before the experiment was run, and no streaming delays were detected. The videos are available as supplementary material to this thesis.

Bartneck, C., & Keijsers, M. (2020). The Morality Of Abusing A Robot. Paladyn - Journal of Behavioral Robotics, 11(1), 271-283. doi:10.1515/pjbr-2020-0017

2.4 Procedure

Prospective participants could select the task in MTurk to read a short description of the study. If they decided to participate, they were directed to a Qualtrics survey page. After informed consent was provided and demographics (age and gender) were assessed, the participants were randomly assigned to one of the two agent conditions (human or robot). They were then given instructions to ensure that their audio playback was working. Within each condition, they watched the videos in a randomised order. After watching each video, the participants provided responses on the dependent measures, before moving to the next video (Figure 2). After the main experiment, the participants filled out the individual differences in the anthropomorphism scale and the two control questions. Then they were thanked for their time, offered a debriefing, and given the reimbursement code which they could use to claim their reward at MTurk. The entire experiment took about 10 min to complete.

3 Results

3.1 Preliminary analyses

3.1.1 Exclusion of participants

Thirty-nine participants were excluded due to reporting that the video material was unrealistic ($n = 34$) and/or reporting that they had already seen the material before ($n = 9$). Of the participants who deemed the material unrealistic, 23 saw the human agent video and 11 saw the robotic agent video. Of the nine participants who reported that they had seen the images before, six were in the robotic agent condition. Participants who were excluded had lower levels of individual tendency to anthropomorphise, M(SD) = 2.31(0.77), than participants who were left in the analysis, M(SD) = 2.63(0.81), $t(75.94) = -2.28$, $p = 0.026$. Excluding those participants thus may have introduced a bias to the results. We discuss the potential biases in more detail in Section 4.1. We chose to report the results for the data set with those participants excluded but ran the same series of tests for the whole data set. If the findings diverged, that is, if a significant effect became insignificant or vice versa, we reported both the results on the full data set as well.

3.1.2 Confound check for an interaction effect between agent and video material

Considering how the abusive behaviours covered a wide range of bullying behaviours, the possibility exists that one or more specific abuses would be considered unacceptable for one agent but not the other. This would be a confound, as the 14 measures are later on collapsed on a single index of abusive behaviour.

Thus, we tested for an interaction between each video and the agent on each of the four measurements. That is, moral acceptability, perceived violence, abusiveness, and the intention to hurt. Note that the objective of this test was explicitly not to look for main effects. Any differences in perceived violence between video A and video B, or in ratings of abusiveness between robotic and human agents, were not considered. Instead, the main points of interest were the interaction effects. For example, whether people thought that the behaviour in video C was way more hurtful than the behaviour in all other videos but only if the agent was a robot. Without such an interaction, the 14 measurements could be aggregated into a single measure of how morally acceptable overall mistreatment of the agent was perceived. The 14 videos thereby became a representative index of abusive behaviour.

For each of the four measurements, a 14 × 2 mixed effects analysis of variance (ANOVA), with the measurement as DV, the video ID as the within-subject factor, and the agent as between-subject factor, was carried out. None of the interaction terms was significant, $\chi^2(13) < 14.80$, ps > 0.320, indicating that it was possible to average the 14 into a single "moral acceptability" measure.

In addition, a confound check was carried out for the two reactive aggression videos (i.e. depicting the agent fighting back). For each of the four measurements, a 2 × 2 mixed effects ANOVA, with the measurement as DV, the video ID as the within-subject factor, and the agent as the between-subject factor, was carried out. None of the interaction terms was significant, $\chi^2(1) < 0.801$, ps > 0.371, indicating that it was possible to average the two videos into a single moral acceptability measure.

3.1.3 Reliability tests

Internal consistency was high for the individual tendency to anthropomorphise scale, $\alpha = 0.83$. The mind attribution scale too had high internal consistency, $\alpha = 0.98$. The scales were therefore deemed reliable [50].

3.1.4 Randomisation tests

Individual tendency to anthropomorphise did not differ between conditions, M(SD) = 2.61(0.69) and 2.64(0.91) for human and robotic agents, respectively, $t(125) = -0.160$, $p =$

0.873. Gender was evenly distributed between the conditions, with the 25 (42.38%) of 59 in the human condition being male and 40 (58.82%) of 68 in the robotic agent condition; $\chi^2(1) = 2.79$, $p = 0.09$. For the full data set, however, gender was not evenly distributed, with 38 (45.24%) of the 82 participants in the human agent condition being male and 51 (62.20%) of 82 being male in the robotic agent condition. As gender was not related to the DV nor the mediator variable, $\chi^2(1) < 0.89$, $p > 0.344$, this imbalance was not considered problematic.

3.1.5 Pearson correlation between acceptability and violence, abusiveness, and intention to harm

Pearson correlation coefficients were calculated between acceptability of robot abuse on one hand, and perceived violence, perceived abusiveness, and intention to harm on the other. The correlation coefficients all exceeded the benchmark for a large effect [48], $\rho > 0.637$, $p < 0.001$. As a result, assessing acceptability of abuse with a single-item measure is not considered problematic for the construct validity.

3.2 Main analyses

To answer our first two research questions, two independent sample t tests are conducted. A 2 × 2 mixed ANOVA with factors "agent" (human versus robot; between participants) and "aggression" (unprovoked, i.e. by the bullies versus reactive, i.e. by the agent; within participants) was considered but would likely have resulted in biased outcomes for the "aggression" main effects, as the unprovoked videos were both more diverse and intense (ranging from verbal abuse to lethal aggression) and higher in number.

The adjusted R^2 (R^2_{adj}) and the Akaike information criterion (AIC) [51] are reported as indication of goodness-of-fit for the significant ANOVA and regression models.

3.2.1 Acceptability of aggression towards agent

Participants in the human agent condition rated the videos as equally acceptable, $M(\text{SD}) = 2.40(0.49)$, as participants in the robot agent condition, $M(\text{SD}) = 2.53(0.80)$; $t(125) = 1.10$, $p = 0.275$. (For the full data set, a marginal effect was found, with abuse of a robotic agent being slightly more acceptable, $M(\text{SD}) = 2.61(0.80)$, than abuse of a human agent, $M(\text{SD}) = 2.41(0.56)$, $t(164) = -1.79$, $p = 0.076$.)

3.2.2 Acceptability of reactive aggression from agent

Participants in the human agent condition rated the reactive aggression as more acceptable, $M(\text{SD}) = 3.74(1.45)$ than participants in the robotic agent condition, $M(\text{SD}) = 2.99(1.33)$, $t(125) = -3.02$, $p = 0.003$; $R^2_{\text{adj}} = 0.06$, AIC = 447.46.

3.2.3 Difference between agents: mediation

In order to answer our third research question, mediation analyses on the significant findings as indicated by the t tests (if any) are conducted. In a mediation analysis, one tries to gain further understanding of a relationship between an independent variable (IV; here: agent) and a DV (here: acceptability of aggression) by including a third variable in the analysis (mediator; here: perceived abuse), which is hypothesised to be related to both the DV and the IV. A mediation model proposes that (part of) the relationship between the IV and the DV is because the IV has an effect on the mediator, which in turn influences the DV. In the current experiment, this would mean that any relationship between agent and the acceptability of aggressive behaviour shown in the video would (partially or only) exist because the agent would have an effect on how abusive the behaviour was perceived to be and how abusive the behaviour was seen to be influenced by the moral acceptability of it.

Mediation analysis can only be performed on a significant relationship between an IV and a DV. Since a significant effect of agent on acceptability had only been established for reactive aggression, acceptability of aggression towards the agent will not be considered for mediation analysis.

For the first step of the mediation analysis, a regression model was specified with acceptability as DV and agent as IV. As found in Section 3.2.2, this relationship was significant, $b = -0.74$, $t(125) = -3.02$, $p = 0.003$, $R^2_{\text{adj}} = 0.06$, AIC = 447.46. In Figure 3, this relationship is shown in the direct effect between agent and moral acceptability.

For the second step, a significant relationship between the independent factor and the mediator has to be established. Abuse was regressed on agent; this relationship, too, was significant, $b = 1.08$, $t(125) = 4.13$, $p < 0.001$. For the final step of the mediation analysis, acceptability was regressed on both agent and abuse. A full mediation occurred, with agent dropping as a predictor

Bartneck, C., & Keijsers, M. (2020). The Morality Of Abusing A Robot. Paladyn - Journal of Behavioral Robotics, 11(1), 271-283. doi:10.1515/pjbr-2020-0017

(b = −0.19, $t(124)$ = −0.86, p = 0.393) and perceived abusiveness taking over as the only significant predictor (b = −0.51, $t(124)$ = −7.22, p < 0.001), R^2_{adj} = 0.33, AIC = 404.91. Sobel's test confirmed the significance of the mediation effect, Z = −3.59, p < 0.001. The difference in AIC scores is 42.55, which indicates that the mediation model fit considerably better than the single IV model [52; AIC differences over three indicate non-equal models]. See Figure 3 for the mediation model.

3.2.4 Cluster analysis

An exploratory cluster analysis was planned on the 14 videos. We expected the videos might cluster on their type of violence (e.g. verbal abuse, physical abuse, and taunting), which would mean the clusters had to be roughly the same for humans and robots. As an initial check for this hypothesis, the 14 videos were plotted on their intention to harm and their violence (Figure 4). The resulting plot did not give reason to suspect any clusters of videos. To confirm, Ward's hierarchical clustering method [53] was employed[4] on both the human and the robotic agent videos. No meaningful clusters could be discerned. It appears that while the videos varied in how abusive they were seen, there were no consistent video "types".

4 Discussion

This experiment compared the differences in acceptability between abuse of a human and abuse of a robotic victim. Markedly, the video materials that were used were both of exceptional quality (making it hard to recognise the robot for the CGI rendering it was) and showed the exact same bullying behaviour to either of the two agents. In addition, the materials covered a wide range of bullying behaviours. As a result, the materials used were both highly realistic and perfectly synchronised except for the agent depicted.

Four research questions were assessed: Is robot abuse seen as more acceptable? Is reactive aggression coming from a human victim seen as more acceptable? Are there any differences between the agents due to perceived abusiveness? And do different abusive behaviours cluster based on their perceived violence and intention to hurt?

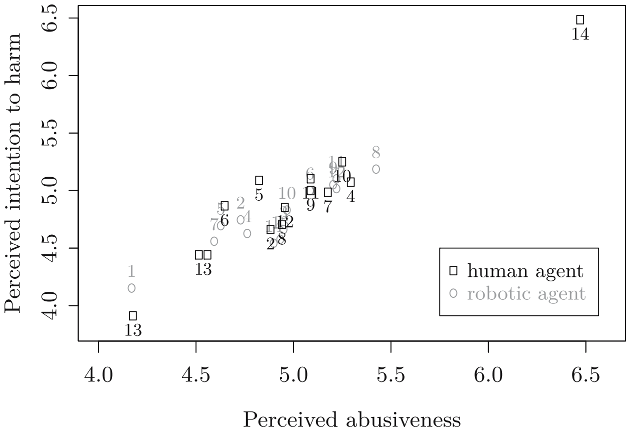


Figure 3: Plot of the 14 videos on perceived intention to hurt and abuse, separated for human and robotic agents. The numbers refer to the video IDs.

First, in line with our predictions, no difference was found in the acceptability of abusive behaviour towards robots or humans. The participants considered mistreating a robot to be as immoral as abusing a human. While this may not automatically mean that the participants consider robots to be equivalent to humans in all respects and in all situations, it does show at least that bullying behaviour is considered immoral, no matter who the victim is.

Two explanations are possible for the similarities between the judgements about the abusive behaviour towards the human and the robot. The participants in our experiment could have considered the robot to be human-like and hence rate the abusive behaviour towards it similar to that towards humans. Alternatively, they could have considered abusive behaviour in itself to be immoral. Sparrow [27] argued that negative behaviour towards robots can be considered immoral because of what it expresses about bullying and about the character of the aggressor. In the reactive aggression condition, however, the participants rated the moral acceptability of abusive behaviour towards a human different from that towards a robot. This would indicate that the similarities in responses observed towards the first 14 videos are more likely to be a result of anthropomorphism of the robot.

It is necessary to point out that the virtual Atlas robot shown has no feelings that it could be hurt nor it could experience pain. The robot's level of sentience or its capacity of feeling pain had not been made explicit at the start of the study, however, so the participants had no way of telling whether the bullying behaviour had any meaning or relevant consequences.

The human and the robot only showed signs of disorientation or malfunctioning in some videos, such as

[4] https://www.statmethods.net/advstats/cluster.html

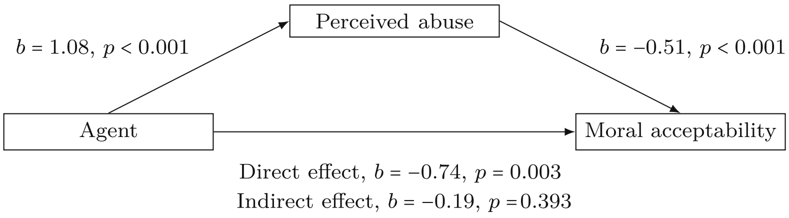


Figure 4: Mediation model showing the full mediation by perceived video on the relationship between agent and moral acceptability of reactive aggression. Note that the default for agent was set at *human*, so that reactive aggression was seen as less acceptable and more abusive for the robotic agent.

stumbling after being spray painted in the face or staggering to the side as an air horn was blown next to their head. Only after 14 videos showing abuse did the agents display reactive aggression. If the participants would have thought that the robot has no emotions or ability to feel pain, then they should have rated the abusive behaviour towards it as less immoral. However, the participants did not.

The second research question was whether reactive aggression from a human would be seen as more acceptable than reactive aggression from a robot. In contrast to the predictions, this was shown to be the case. The robot fighting back was considered less acceptable than the human fighting back. The third research question was answered by means of a mediation analysis. This showed that the difference in moral acceptability of reactive aggression was entirely due to participants perceiving the robot's response as more abusive than the human's response. We need to point out again the fact that the acts of responsive aggression were identical. What could have caused this asymmetry?

We speculate that robots could be perceived to deserve protection from harm to the same extent as humans but are not perceived to have the same right of self-defence. To our knowledge, there are two HRI papers that relate to these findings. Kahn et al. [15] had children of various ages interact with a Robovie humanoid robot before it was locked away in a closet. Robovie protested against this treatment. The children were interviewed about a range of topics, including the robot's moral standing. If we only consider the oldest age group (the 15-year-olds), an interesting pattern emerged. Slightly more than half of the 15-year-olds thought it was wrong to hurt the robot by locking it away or eventually crushing it when it would be no longer needed. The vast majority, however, did not think the robot should be paid for a hard day's work or be granted the right to vote; and less than 1 in 10 thought the concept of owning and selling the robot was wrong. This pattern – a right to be protected from harm but no right to autonomy – is surprisingly similar to what was found in our study. It mirrors the ethical view many have towards animal rights. While animals can be considered property and can even be killed, mistreatment is not allowed. Animal rights has therefore been proposed as a template for robot rights [54].

A different perspective could be offered by a study on the trolley dilemma [33]. In this moral dilemma, a trolley is rushing down the track at great speed and will hit and kill four people if not sidetracked to a route where it will kill only one person. People have to choose between not taking any action, and thus indirectly being responsible for the death of four people, or taking action and being directly responsible for the death of one person. In spite of the net saving of three lives when taking action, most people find taking action harder than not taking any action. However, robots were more strongly expected to make a rational choice and more strongly blamed if they go with the emotional solution. On the contrary, humans were blamed more if they chose to divert the train. In this light, the robot's reactive aggression could be considered more wrong as we expect robots to be more rational and less affected by emotion when choosing to take action, while emotion is expected to play a role in human moral decision-making.

A third possible explanation can be found in how intimidating the behaviour was seen. A robot fighting back might be considered as a threat as robots are often portrayed in public media as a potential threat. The trope is that robots raise up against their masters and enslave humanity [55,56].

Finally, the fourth and last research question concerned whether the abusive behaviours clustered based on perceived violence and intention to hurt. For example, there might have been clusters for verbal abuse, lethal abuse, and (physical) taunting abuse. In addition, these might have been different for the human and the robotic agent, as people could have reasoned that verbal abuse of a human has a higher intention to harm than verbal abuse of a robot. However, our analyses could show no such clusters, agent-specific or overall.

Bartneck, C., & Keijsers, M. (2020). The Morality Of Abusing A Robot. Paladyn - Journal of Behavioral Robotics, 11(1), 271-283. doi:10.1515/pjbr-2020-0017

4.1 Limitations

This experiment has a few limitations that need to be noted.

First, the videos showing the human actor were used to motion capture the movement of the digital Atlas robot. For this purpose, the human actor moved in a rigid, mechanical way, which slightly deviates from the natural human movement. Still, humans need to move in plausible biological ways and therefore the actor's movement can still clearly be identified as that of a human. If the human actor had moved in a different way than the Atlas robot, then this would have introduced another possible bias. We believe that the movements of the actor were sufficiently plausible movements for a human.

A second limitation of this study is that the Corridor Digital company added some small special effects to the robot video, in particular in the video in which a bully shoots a handgun at the robot. In the robot video there is additional nozzle fire, smoke, and impact indicators. All these effects are subtle and the drama of the scene certainly outweighs these minor differences. The human video also contains light effects that indicate gunfire. We believe that these slight differences do not significantly impact the similarities of the videos. Most videos were identical between the human and the robot conditions.

We excluded participants who considered the videos as unrealistic from our statistical analyses. This might have introduced a small bias since these participants also had a slightly different tendency to anthropomorphise. Still including them would have also introduced another bias, namely, that of participants who did not suspend their disbelief. We believe that the latter would have been the stronger bias and hence our decision to exclude the participants was the better choice. To have a better insight into how the exclusion may have influenced the results, we ran the same analyses once more with the complete data set. No new significant results emerged and no previously significant results turned insignificant.

Acknowledgements: The authors would like to acknowledge Jake Watson and Sam Gorski from Corridor Digital who made the stimuli for this experiment available.

References

[1] D. Mosbergen, "Good Job, America. You Killed hitchBOT," *Huffpost*, 2015, https://www.huffpost.com/entry/hitchbot-destroyed-philadelphia_n_55bf24cde4b0b23e3ce32a67.

[2] D. Brscić, H. Kidokoro, Y. Suehiro, and T. Kanda, "Escaping from children's abuse of social robots," in: *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction, HRI '15*, ACM, New York, NY, USA, 2015, pp. 59–66, DOI: 10.1145/2696454.2696468.

[3] J. Vincent, "A drunk man was arrested for knocking over silicon Valley's crime-fighting robot," *Verge*, 2017, https://www.theverge.com/2017/4/26/15432280/securityrobot-knocked-over-drunk-man-knightscope-k5-mountainview.

[4] M. Keijsers, H. Kazmi, F. Eyssel, and C. Bartneck, "Teaching robots a lesson: determinants of robot punishment," *Int. J. Soc. Robotics*, 2019, DOI: 10.1007/s12369-019-00608-w.

[5] M. Keijsers and C. Bartneck, "Mindless robots get bullied," in: *ACM/IEEE International Conference on Human Robot Interaction*, ACM Press, Chicago, 2018, pp. 205–214.

[6] T. Nomura, T. Kanda, H. Kidokoro, Y. Suehiro, and S. Yamada, "Why do children abuse robots?" *Interact. Stud.*, vol. 17, no. 3, pp. 347–369, 2017.

[7] A. J. Nederhof, "Methods of coping with social desirability bias: a review," *Eur. J. Soc. Psychol.*, vol. 15, no. 3, pp. 263–280, 1985.

[8] D. Ritter and M. Eslea, "Hot sauce, toy guns, and graffiti: a critical account of current laboratory aggression paradigms," *Aggressive Behavior: Off. J. Int. Soc. Res. Aggression*, vol. 31, no. 5, pp. 407–419, 2005.

[9] C. Bartneck, M. Verbunt, O. Mubin, and A. A. Mahmud, "To kill a mockingbird robot," in: *2nd ACM/IEEE International Conference on Human-Robot Interaction*, ACM Press, 2007, pp. 81–87, 1179031.

[10] Z. Carlson, L. Lemmon, M. Higgins, D. Frank, and D. Feil-Seifer, "This robot stinks! Differences between perceived mistreatment of robot and computer partners," preprint arXiv:1711.00561, 2017.

[11] H. Lucas, J. Poston, N. Yocum, Z. Carlson, and D. Feil-Seifer, "Too big to be mistreated? Examining the role of robot size on perceptions of mistreatment," in: *2016 25th IEEE International Symposium on Robot and Human Interactive Communication* (*RO-MAN*), IEEE, 2016, pp. 1071–1076.

[12] X. Z. Tan, M. Vázquez, E. J. Carter, C. G. Morales, and A. Steinfeld, "Inducing bystander interventions during robot abuse with social mechanisms," in: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, ACM, 2018, pp. 169–177.

[13] A. M. Rosenthal-von der Pütten, N. C. Krämer, L. Hoffmann, S. Sobieraj, and S. C. Eimler, "An experimental study on emotional reactions towards a robot," *Int. J. Soc. Robot.*, vol. 5, no. 1, pp. 17–34, 2013.

[14] A. M. Rosenthal-von der Pütten, F. P. Schulte, S. C. Eimler, S. Sobieraj, L. Hoffmann, S. Maderwald, et al., "Investigations on empathy towards humans and robots using fMRI," *Computers Hum. Behav.*, vol. 33, pp. 201–212, 2014.

[15] P. H. Kahn Jr, T. Kanda, H. Ishiguro, N. G. Freier, R. L. Severson, B. T. Gill, et al., "'Robovie, you'll have to go into the closet now': Children's social and moral relationships with a humanoid robot," *Dev. Psychol.*, vol. 48, no. 2, 303–314, 2012.

[16] V. Gazzola, G. Rizzolatti, B. Wicker, and C. Keysers, "The anthropomorphic brain: the mirror neuron system responds to human and robotic actions," *Neuroimage*, vol. 35, no. 4, pp. 1674–1684, 2007.

[17] S. Krach, F. Hegel, B. Wrede, G. Sagerer, F. Binkofski, and T. Kircher, "Can machines think? Interaction and perspective

taking with robots investigated via fMRI," *PLoS One*, vol. 3, no. 7, e2597, 2008.

[18] T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro, "A two-month field trial in an elementary school for long-term human-robot interaction," *IEEE Trans. Robot.*, vol. 23, no. 5, pp. 962–971, 2007.

[19] C. Bartneck, M. v. d. Hoek, O. Mubin, and A. A. Mahmud, "'Daisy, Daisy, Give me your answer do!' – Switching off a robot," in: *2nd ACM/IEEE International Conference on Human-Robot Interaction*, ACM Press, 2007, pp. 217–222.

[20] C. Bartneck, J. Reichenbach, and J. Carpenter, "The carrot and the stick – The role of praise and punishment in human-robot interaction," *Interact. Stud. – Soc. Behav. Commun. Biol. Artif. Syst.*, vol. 9, no. 2, pp. 179–203, 2008.

[21] M. Slater, A. Antley, A. Davison, D. Swapp, C. Guger, C. Barker, et al., "A virtual reprise of the Stanley Milgram obedience experiments," *PLoS One*, vol. 1, no. 1, e39, 2006.

[22] K. Darling, "Extending legal rights to social robots," in: *We Robot Conference*, University of Miami, Miami, USA, 2012, pp. 1–24.

[23] P. Salvini, G. Ciaravella, W. Yu, G. Ferri, A. Manzi, B. Mazzolai, C. Laschi, et al., "How safe are service robots in urban environments? Bullying a robot," in: *RO-MAN, 2010 IEEE*, Viareggio, Italy, 2010, pp. 1–7.

[24] D. Brscić, H. Kidokoro, Y. Suehiro, and T. Kanda, "Escaping from children's abuse of social robots," in: *Proceedings of the 10th annual ACM/IEEE International Conference on Human-Robot Interaction*, ACM/IEEE, Portland, USA, 2015, pp. 59–66.

[25] C. Bartneck and J. Hu, "Exploring the abuse of robots," *Interact. Stud.*, vol. 9, no. 3, pp. 415–433, 2008.

[26] A. De Angeli, S. Brahnam, P. Wallis, and A. Dix, "Misuse and abuse of interactive technologies," in: *CHI'06 Extended Abstracts on Human Factors in Computing Systems*, ACM, Montreal, Canada, 2006, pp. 1647–1650.

[27] R. Sparrow, "Robots, rape, and representation," *Int. J. Soc. Robot.*, vol. 9, no. 4, pp. 465–477, 2017, DOI: 10.1007/s12369-017-0413-z.

[28] B. Whitby, "Sometimes it's hard to be a robot: a call for action on the ethics of abusing artificial agents," *Interact. Computers*, vol. 20, no. 3, pp. 326–333, 2008.

[29] K. Richardson, "The asymmetrical 'relationship': parallels between prostitution and the development of sex robots," *ACM SIGCAS Computers Soc.*, vol. 45, no. 3, pp. 290–293, 2016.

[30] K. Darling, "'Who's Johnny?'Anthropomorphic framing in human-robot interaction, integration, and policy," in: P. Lin, K. Abney, and R. Jenkins, Eds., *Robot Ethics 2.0: from Autonomous Cars to Artificial Intelligence*, ch. 12, Oxford University Press, Oxford, 2015.

[31] E. B. Sandoval, J. Brandstetter, and C. Bartneck, "Can a robot bribe a human? The measurement of the dark side of reciprocity in human robot interaction," in: *11th ACM/IEEE International Conference on Human-Robot Interaction*, IEEE, Christchurch, 2016, pp. 117–124.

[32] B. Reeves and C. Nass, *The Media Equation*, CSLI Publications and Cambridge University Press, Cambridge, 1996.

[33] B. F. Malle, M. Scheutz, T. Arnold, J. Voiklis, and C. Cusimano, "Sacrifice one for the good of many?: People apply different moral norms to human and robot agents," in: *Proceedings of the 10th annual ACM/IEEE International Conference on Human-Robot Interaction*, ACM, 2015, pp. 117–124.

[34] J. Li, "The benefit of being physically present: a survey of experimental works comparing copresent robots, telepresent robots and virtual agents," *Int. J. Human-Computer Stud.*, vol. 77, pp. 23–37, 2015.

[35] J. Reichenbach, C. Bartneck, and J. Carpenter, "Well done, Robot! The importance of praise and presence in human-robot collaboration," in: *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication* (*RO-MAN 2006*), IEEE, Hatfield, UK, 2006, pp. 86–90.

[36] A. Powers, S. Kiesler, S. Fussell, and C. Torrey, "Comparing a computer agent with a humanoid robot," in: *Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction* (*HRI 2007*), ACM/IEEE, Arlington, USA, 2007, 145–152.

[37] R. Wullenkord, M. R. Fraune, F. Eyssel, and S. Sabanović, "Getting in Touch: How imagined, actual, and physical contact affect evaluations of robots," in: *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication* (*RO-MAN 2016*), IEEE, New York, USA, 2016, pp. 980–985.

[38] S. Thellman, A. Silvervarg, A. Gulz, and T. Ziemke, "Physical vs. virtual agent embodiment and effects on social interaction," in: *Intelligent Virtual Agents: 16th International Conference* (*IVA 2016*), Springer International Publishing, Los Angeles, USA, 2016, pp. 412–415.

[39] K. M. Lee, "Presence, explicated," *Commun. Theory*, vol. 14, no. 1, pp. 27–50, 2004.

[40] R. Sparrow, "Kicking a robot dog," in: *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction* (*HRI 2016*), IEEE, Christchurch, 2016, pp. 229–229.

[41] B. Dynamics, *Testing Robustness*, https://youtu.be/aFuA50H9uek, 2018.

[42] C. Nass, J. Steuer, and E. R. Tauber, "Computers are social actors," in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (*CHI '94*), Association for Computing Machinery, New York, NY, USA, 1994, pp. 72–78, DOI: 10.1145/191666.191703.

[43] C. Bartneck, A. Duenser, E. Moltchanova, and K. Zawieska, "Comparing the similarity of responses received from studies in Amazon's Mechanical Turk to studies conducted online and with direct recruitment," *PLoS One*, vol. 10, no. 4, e0121595, 2015.

[44] D. J. Simons and C. F. Chabris, "Common (mis)beliefs about memory: a replication and comparison of telephone and Mechanical Turk survey methods," *PLoS One*, vol. 7, no. 12, e51876, 2012, DOI: 10.1371/journal.pone.0051876.

[45] M. Buhrmester, T. Kwang, and S. D. Gosling, "Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?" *Perspect. Psychol. Sci.*, vol. 6, no. 11, pp. 3–5, 2011.

[46] B. F. Malle, P. Bello, and M. Scheutz, "Requirements for an artificial agent with norm competence," in: *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (*AIES '19*), ACM, New York, NY, USA, 2019, pp. 21–27, DOI: 10.1145/3306618.3314252.

[47] R. W. Lissitz and S. B. Green, "Effect of the number of scale points on reliability: A Monte Carlo approach," *J. Appl. Psychol.*, vol. 60, no. 1, pp. 10–13, 1975.

[48] J. Cohen, "A power primer," *Psychol. Bull.*, vol. 112, no. 1, pp. 155–159, 1992.

[49] A. Waytz, J. Cacioppo, and N. Epley, "Who sees human? The stability and importance of individual differences in anthropomorphism," *Perspect. Psychol. Sci.*, vol. 5, no. 3, pp. 219–232, 2010.

[50] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, no. 3, pp. 297–334, 1951.

[51] H. Bozdogan, "Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical extensions," *Psychometrika*, vol. 52, no. 3, pp. 345–370, 1987, DOI: 10.1007/BF02294361.

[52] Anonymous, "Analysis of Longitudinal Data," *Technometrics*, vol. 45, no. 2, pp. 181–181, 2003, doi: 10.1198/tech.2003.s147.

[53] J. H. Ward Jr, "Hierarchical grouping to optimize an objective function," *J. Am. Stat. Assoc.*, vol. 58, no. 301, pp. 236–244, 1963.

[54] D. J. Calverley, "Android science and animal rights, does an analogy exist?" *Connect. Sci.*, vol. 18, no. 4, pp. 403–417, 2006.

[55] J. Zlotowski, K. Yogeeswaran, and C. Bartneck, "Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources," *Int. J. Human-Computer Stud.*, vol. 100, pp. 48–54, 2017, DOI: 10.1016/j.ijhcs.2016.12.008.

[56] C. Bartneck, "Robots in the theatre and the media," in *Design and Semantics of Form and Movement* (*DeSForM2013*), Philips, 2013, pp. 64–70, DOI: 10.1016/j.ijhcs.2016.12.008.