International Journal of Social Robotics
https://doi.org/10.1007/s12369-020-00663-8

I, Robot: How Human Appearance and Mind Attribution Relate to the Perceived Danger of Robots

Barbara C. N. Müller[1] · Xin Gao[2] · Sari R. R. Nijssen[1] · Tom G. E. Damen[3]

Accepted: 28 May 2020

Abstract
Social robots become increasingly human-like in appearance and behaviour. However, a large body of research shows that these robots tend to elicit negative feelings of eeriness, danger, and threat. In the present study, we explored whether and how human-like appearance and mind-attribution contribute to these negative feelings and clarified possible underlying mechanisms. Participants were presented with pictures of mechanical, humanoid, and android robots, and physical anthropomorphism (Studies 1–3), attribution of mind perception of agency and experience (Studies 2 and 3), threat to human–machine distinctiveness, and damage to humans and their identity were assessed for all three robot types. Replicating earlier research, human–machine distinctiveness mediated the influence of anthropomorphic appearance on the perceived damage for humans and their identity, and this mediation was due to anthropomorphic appearance of the robot. Perceived agency and experience did not show similar mediating effects on human–machine distinctiveness, but a positive relation with perceived damage for humans and their identity. Possible explanations are discussed.

Keywords Human/robot interaction · Uncanny valley · Mind perception · Need for distinctiveness

1 I, Robot: How Human Appearance and Mind Attribution Relate to the Perceived Danger of Robots

Watching the movie '*Ex Machina'*, you quickly perceive Ava, the android main character of the movie, as a real human with emotions and feelings. Anthropomorphising Ava in this way, that is to ascribe human-like characteristics and/or intentions to non-human agents, is a fundamental human process that spontaneously happens and increases our social connection with non-human agents [1]. Although highly evolved robots seem a vision of the future, we already interact with artificial intelligent agents on a regular basis (e.g., Siri, Apple's speaking assistant, Amazon's Alexa, or CaMeLi, an avatar designed to help elderly in daily life). Developments in robot technology are proceeding rapidly: 'Social robots', i.e., robots that are designed to interact and communicate with people [2], feature increasingly more human-like appearances and behaviour. While, these technical developments are especially interesting when it comes to maintaining and improving our quality of life, for example in health care or education, a large body of research also shows that social robots tend to elicit negative feelings of eeriness, danger, and threat [3–7]. In the present study, we investigated the factors that elicit these negative feelings and clarified possible underlying mechanisms, including the extent to which robots look human-like and the extent to which they are attributed with a mind.

1.1 Anthropomorphism

Anthropomorphism involves "going beyond behavioral descriptions of imagined or observable actions (e.g., the dog is affectionate) to represent an agent's mental or physical characteristics using humanlike descriptors (e.g., the dog loves me)" [1, page 865]. Anthropomorphism for non-human agents can be elicited in two ways: First, by

✉ Barbara C. N. Müller
B.Muller@bsi.ru.nl

[1] Behavioural Science Institute, Radboud University Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands

[2] Marketing and Consumer Behaviour Group, Wageningen University, P.O. Box 8130, 6700 EW Wageningen, The Netherlands

[3] Department of Psychology, University Utrecht, Heidelberglaan 1, 3584 CS Utrecht, The Netherlands

Published online: 18 June 2020

increasing the physical appearance with humans [8–10]. Second, through attributions that ascribe agency, affect, or intentionality to the non-human [11]. Interestingly, humans spontaneously anthropomorphise non-human agents [1]. Initially introduced to describe the appearance of religious agents and gods [Hume, 1757, in 1], the term is now used to describe human characteristics towards animals or plants [12], objects and technical devices [13], and even geometric shapes [14]. In fact, neuroscientific research has demonstrated that similar brain regions are activated when participants attribute mental states to non-human agents as when attributing mental states to other humans [15–17]. In a predictive coding framework [18, 19], which suggests that the brain continuously produce hypotheses that predict sensory input, anthropomorphism makes sense: when something looks like a human or moves or acts in a human-like way, it is more likely that your interaction with this agent will be efficient and smooth if you treat it as another human-being.

Research has shown that when we anthropomorphise non-biological agents, this has profound effects on how we interact with these agents: it leads to more interpersonal closeness [20], increased moral care [21], and smoother interactions [11, 22]. In addition, people rate robots acting playfully as being more extroverted and outgoing than robots acting seriously [23, 24]. Moreover, avatars are judged to be more trustworthy, competent, sensitive, and warm when they are anthropomorphised [25], and similar stereotypes are applied to robots than to humans [26].

1.2 Negative Effects of Anthropomorphism

However, perceptual similarity with humans can also elicit negative feelings towards robots: Robots whose physical appearance closely (but not perfectly) resembles human beings often evoke negative feelings, a phenomenon referred to as the *uncanny valley* [4]. The uncanny valley hypothesis states that more realistic artificial agents elicit more positive reactions until they are very close (but not close enough) to the human ideal. This dimension of "realism", however, is not limited to realistic looks/appearances, but also realistic motor behaviours, such as imitation [4]. Research showed that human-like appearance or motion can be responsible for the emergence of the uncanny valley [27, 28]. Furthermore, large inter-individual differences in the emergence of the uncanny valley have been found, suggesting that people can be more or be less sensitive towards negative feelings in response to a robot [5]. Often, the concept of prediction error has been used to explain the uncanny valley [29]: when something looks like a human or moves or acts in a human-like way, it is more likely that your interaction with this agent will be efficient and smooth if you treat it as another human-being. However, when a robot that looks extremely human-like moves in a mechanical way, our predictions are violated since their strong anthropomorphic appearance led us to expect them to follow biological movement patterns. Consequently, this high prediction error leads to a negative feeling of unease or fear.

Interestingly, an additional explanation of the uncanny valley has been proposed recently, with empirical studies suggesting that this negativity for human-like robots can be explained by a violation of the need for distinctiveness [30, 31]. As humans, we feel unique and distinct by understanding how our own group differs from another group [32, 33]. However, when this feeling of uniqueness, and with it our intergroup boundaries, disappears, we feel threatened. Applied to robots, correlational research showed that higher feelings of eeriness and a decrease in felt warmth towards robots was positively related to whether participants perceived robots and humans as similar categories [31]. In addition, it was causally demonstrated that too much perceived similarity between robots and humans undermines people's ideas about human uniqueness, and this subsequently leads people to perceive robots as more threatening and potentially damaging entities [30].

In line with this assumption, looking at manipulation of anthropomorphism, a recent study on human/avatar interactions found evidence for a so-called "uncanny valley of the mind" [6]. In their study, participants watched interactions involving emotional responses between two digital characters, which were presented as either human-controlled versus computer-controlled, and scripted versus autonomous. Their results showed that levels of eeriness rose especially when participants thought the digital agents were autonomous artifical intelligences. These findings support the notion that attributions of a mind might lead to a decrease in human/machine distinctiveness, subsequently leading to a negative feelings and perceived damage to someone's identity.

A key feature by which we distinguish between humans and non-humans relates to mind attribution. Specifically, we attribute the minds of potential other agents along the lines of two constructs: experience and agency [34]. Experience involves capacities to feel emotions, such as the ability to experience hunger or pleasure, whereas agency relates to capacities of being an autonomous agent, such as self-control and thought. Especially experience is seen as a unique human trait, as people ascribe a medium amount of agency but no experience towards robots [34]. The question however is whether mind attribution influences the need for distinctiveness.

In three studies, we aimed to replicate the link between robot-human similarity and threat-perceptions using a variety of stimuli (Studies 1-3). Furthermore, we aimed to extend the literature in this domain by investigating whether distinctiveness-negativity is evoked by perceived physical similarity, by similarity in mind attribution, or both (Study 2 and Study 3; see Fig. 1 for an overview of the mediation

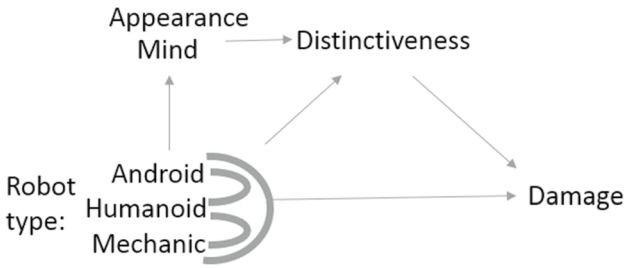

Fig. 1 Overview of the tested mediation model

model). Similar to earlier research [30], participants were presented with pictures of mechanical, humanoid, and android robots, and physical anthropomorphism, mind attribution of agency and experience (Study 2 and Study 3), threat to the human–machine distinctiveness, and damage to humans and their identity were assessed for all three robot types. We expected that human–machine distinctiveness mediated the influence of robot type on the perceived damage for humans and their identity, and that this mediation would be due to (a) the anthropomorphic appearance of the robot, and (b) perceived experience and perceived agency from the robot.

2 Study 1

2.1 Methods

2.1.1 Participants and Design

Fifty-five participants (44 females, 11 males, $M_{age} = 19.50$ years, $SD_{age} = 1.40$, age range 17–23 years) completed the experiment in exchange for course credits. The experiment had a 3 (Robot type: mechanical robot vs. humanoid robot vs. android robot) within-subjects design, with damage to humans and their identity as dependent variable, and physical anthropomorphism and threat to the human–machine distinctiveness as mediators. Data was acquired online using Inquisit 4 [35]. Via an online participant pool from Radboud University, participants could sign up for the study, and received a link to complete the study online. Before participants could start with the experiment, they were asked to ensure they would not be interrupted for the duration of the experiment (approximately 30 min).

2.1.2 Procedure and Materials

Participants were instructed that they had to evaluate different types of robots: Participants received a self-paced evaluation task in which they indicated their attitudes towards three different categories of robots varying in human-like appearance; participants either saw mechanical, humanoid, or android robots. While mechanical robots were clearly machines (no legs, no facial features), humanoid robots were more humanlike by having legs, arms, a torso, and a head with a face. Still, they also possess clear similarities with a machine (e.g., no hair or skin). Android robots were very high in humanlike-ness and difficult to distinguish from real humans. All stimuli and measures used were derived from earlier research [30]. Four different robots were used for each robot category, resulting in 12 robot evaluations.[1]

Participants had to rate each of the 12 robots on (1) physical anthropomorphism; (2) threat to human–machine distinctiveness; and (3) damage to humans and human identity. Firstly, physical anthropomorphism was assessed using a three-item scale (e.g., "I could easily mistake the robot for a real person"; Cronbach's $\alpha_{\text{mechanical robot}} = .935$; Cronbach's $\alpha_{\text{humanoid robot}} = .887$; Cronbach's $\alpha_{\text{android robot}} = .800$).[2] Secondly, threat to the human–machine distinctiveness was measured using a three-item scale (e.g., "Looking this kind of robot I ask myself what the differences are between robots and humans"; Cronbach's $\alpha_{\text{mechanical robot}} = .948$; Cronbach's $\alpha_{\text{humanoid robot}} = .957$; Cronbach's $\alpha_{\text{android robot}} = .963$). Thirdly, damage to humans and human identity was assessed using a four items scale (e.g., "I get the feeling that the robot could damage relations between people"; Cronbach's $\alpha_{\text{mechanical robot}} = .880$; Cronbach's $\alpha_{\text{humanoid robot}} = .895$; Cronbach's $\alpha_{\text{android robot}} = .897$). For all three questionnaires, participants could answer on a 7-point Likert scale (1 = 'totally not agree' to 7 = 'totally agree'). Items were presented along with each robot photo respectively, and robot photos were presented in a random order. After completing the questionnaires, participants reported their age and gender, and were thanked for their participation.

2.2 Results and Discussion

Within-participant mediation analyses were conducted using MEMORE package in SPSS with 1000 bootstrap samples [36]. In this analysis conclusions on mediation are based upon the correlation of the score-difference between conditions of the dependent variable and the score-difference between conditions of the proposed mediator. Since each participant responded to all three conditions, and we did not have specific hypotheses on certain comparisons, we

[1] Please contact the corresponding author for stimuli examples.

[2] Three extra questions were included about robotic appearance (e.g., "The robot looks like a robot") as in previous work [30]. However, as our hypotheses concerned humanlike appearance, we did not include these questions in the analyses.

Table 1 B's and 99% confidence intervals (between brackets) of the mediation analyses of Study 1, as a function of comparison (mechanical vs. humanoid; mechanical vs. android; humanoid vs. android)

Study 1 IV: Robot DV: Damage	Mediation effect	Mechanical versus android	Mechanical versus humanoid	Humanoid versus android
Mediator: Distinctiveness	Indirect effect	2.108, [.84, 3.42]	.483, [.20, .91]	1.215, [.34, 2.18]
	Direct effect	− .010, [− 1.38, 1.36]	.058, [− .26, .38]	.342, [− .59, 1.27]
	Total effect	2.098, [1.52, 2.68]	.541, [.24, .84]	1.557, [1.08, 2.04]
Study 1 IV: Robot DV: Distinctiveness				
Mediator: Appearance	Indirect effect	3.708, [2.83, 6.48]	.788, [.40, 1.31]	2.371, [1.39, 3.40]
	Direct effect	.155, [− 1.08, 1.39]	.187, [− .27, .65]	.517, [− .64, 1.68]
	Total effect	3.863, [3.31, 4.42]	.975, [.57, 1.38]	2.888, [2.34, 3.44]

Table 2 Descriptive statistics (Means, SD) of Study 2 for all variables (N = 55)

Type	Appearance	Distinctiveness	Damage
Mean			
Android	5.57	5.76	4.12
Humanoid	2.51	2.87	2.56
Mechanic	1.67	1.89	2.02
SD			
Android	0.970	1.07	1.50
Humanoid	1.05	1.30	1.13
Mechanic	0.674	0.799	0.918

reported all possible comparisons of three robot types (i.e., mechanical vs. humanoid, mechanical vs. android, and humanoid vs. android) while using Bonferroni correction to control Type one error (here we used 99% confidence interval). All mediation statistics (B and 99% CI) are depicted in Table 1. An overview about descriptive statistics can be found in Table 2.

2.2.1 Mediation Effect of Distinctiveness on the Relation Between Robot Types and Damage to Humans and Their Identity

We conducted a mediation analysis on the relation between robot types as IV and damage to humans and their identity as DV, with the human–machine distinctiveness of the different robot types as a proposed mediator. The model indicated significant indirect effects and non-significant direct effects for all three comparisons (mechanical with android, mechanical with humanoid, and humanoid with android robot). The relation between robot-types and damage to humans and their identity was therefore fully mediated by the human–machine distinctiveness of the different robot types.

2.2.2 Mediation Effect of Anthropomorphic Appearance on the Relation Between Robot Type and Distinctiveness

We conducted a mediation analysis on the relation between robot types as IV and human–machine distinctiveness as DV, with the anthropomorphic appearance of the different robot types as a proposed mediator. The model indicated significant indirect effects and non-significant direct effects for all three comparisons. The relation between robot-types and human–machine distinctiveness was therefore fully mediated by the anthropomorphic appearance of the different robot types.

Our findings of Study 1 successfully replicate earlier research on the relationship between anthropomorphic appearance, human–machine distinctiveness, and perceived threat to humans and their identity [30]: Human–machine distinctiveness mediated the influence of robot type on the perceived damage for humans and their identity, and this mediation was due to anthropomorphic appearance of the robot. Thus, these results support the notion that too much physical appearance leads to negative perceptions of robots due to a decrease in distinctiveness [30].

In our following studies, we add the concept of mind perception to our model [34, 37], a fundamental construct central for us to understand ourselves as living beings. The mind perception literature distinguishes between two essential constructs: agency and experience [34]. Both agency and experience are critical components for being human, and how we are distinct from robots, avatars, and other non-living objects. Our aim in Study 2 was to replicate findings from Study 1, and additionally, to investigate whether agency and experience mediate the relation between robot-types and human machine distinctiveness.

Table 3 B's and 99% confidence intervals (between brackets) of the mediation analyses of Study 2, as a function of comparison (mechanical vs. humanoid; mechanical vs. android; humanoid vs. android)

Study 2 IV: Robot DV: Damage	Mediation effect	Mechanical versus android	Mechanical versus humanoid	Humanoid versus android
Mediator: Distinctiveness	Indirect effect	2.310, [1.46, 3.25]	.472, [.18, .79]	1.354, [.51, 2.34]
	Direct effect	− .250, [− 1.44, .94]	.159, [− .21, .52]	.075, [− .78, .93]
	Total effect	2.059, [1.51, 2.61]	.630, [.36, .90]	1.429, [.96, 1.90]
Study 2 IV: Robot DV: Distinctiveness				
Mediator: Appearance	Indirect effect	3.902, [1.82, 5.12]	.437, [.15, .80]	2.701, [1.73, 3.71]
	Direct effect	− .137, [− 1.75, 1.47]	.547, [.14, .96]	.079, [− .99, 1.15]
	Total effect	3.765, [3.17, 4.36]	.985, [.68, 1.29]	2.781, [2.27, 3.29]
Study 2 IV: Robot DV: Distinctiveness				
Mediator: Experience	Indirect effect	1.221, [− .01, 2.26]	.250, [− .10, .60]	.486, [− .35, .97]
	Direct effect	2.544, [1.33, 3.76]	.736, [.30, 1.17]	2.295, [1.55, 3.04]
	Total effect	3.765, [3.17, 4.36]	.985, [.68, 1.29]	2.781, [2.27, 3.29]
Study 2 IV: Robot DV: Distinctiveness				
Mediator: Agency	Indirect effect	.724, [− .05, 1.54]	.282, [− .06, .67]	.469, [.03, .85]
	Direct effect	3.042, [2.03, 4.05]	.703, [.25, 1.16]	2.312, [1.73, 2.89]
	Total effect	3.765, [3.17, 4.36]	.985, [.68, 1.29]	2.781, [2.27, 3.29]

3 Study 2

3.1 Methods

3.1.1 Participants

Sixty participants (52 females, 5 males, 3 unknown, $M_{age} = 19.30$ years, $SD_{age} = 1.90$, age range 17–24 years) completed the experiment for course credits. The experiment had a 3 (Robot type: mechanical robot vs. humanoid robot vs. android robot) within-subjects design, with damage to humans and their identity as dependent variable, and physical anthropomorphism, mind attribution, and threat to the human–machine distinctiveness as mediators. Data was acquired similar to Study 1.

3.1.2 Procedure and Materials

The same procedure and stimuli from Study 1 was used. However, in Study 2, the mind attribution scale was added (Gray et al., 2007), and subsequently, the other variables (physical anthropomorphism: Cronbach's $\alpha_{\text{mechanical robot}} = .884$; Cronbach's $\alpha_{\text{humanoid robot}} = .832$; Cronbach's $\alpha_{\text{android robot}} = .785$; threat to human–machine distinctiveness: Cronbach's $\alpha_{\text{mechanical robot}} = .948$; Cronbach's $\alpha_{\text{humanoid robot}} = .957$; Cronbach's $\alpha_{\text{android robot}} = .974$; damage to humans and their identity: Cronbach's $\alpha_{\text{mechanical robot}} = .932$; Cronbach's $\alpha_{\text{humanoid robot}} = .937$; Cronbach's $\alpha_{\text{android robot}} = .918$) were assessed. Participants had to rate 18 adjectives on whether they think the robot can experience certain capacities connected to agency (e.g., thought, self-control; 7 items; Cronbach's $\alpha_{\text{mechanical robot}} = .841$; Cronbach's $\alpha_{\text{humanoid robot}} = .884$; Cronbach's $\alpha_{\text{android robot}} = .898$) and experience (feelings such as anger, joy; 11 items; Cronbach's $\alpha_{\text{mechanical robot}} = .943$; Cronbach's $\alpha_{\text{humanoid robot}} = .967$; Cronbach's $\alpha_{\text{android robot}} = .978$). A mean score for all dependent variables was calculated before data analyses.

3.2 Results and Discussion

All mediation statistics (B and 99% CI) are depicted in Table 3. An overview about descriptive statistics can be found in Table 4.

3.2.1 Mediation Effect of Distinctiveness on the Relation Between Robot Types and Damage to Humans and Their Identity

We conducted the same mediation analysis as in Study 1: We treated robot types as IV and damage to humans and their identity as DV, and human–machine distinctiveness of the different robot types as a proposed mediator. The same

Table 4 Descriptive statistics (Means, SD) of Study 2 for all variables (N = 60)

Type	Appearance	Experience	Agency	Distinctiveness	Damage
Mean					
Android	5.35	3.95	4.50	5.14	4.03
Humanoid	2.11	2.58	3.60	2.36	2.60
Mechanical	1.25	1.57	2.30	1.38	1.97
SD					
Android	0.910	1.86	1.48	1.37	1.60
Humanoid	0.859	1.30	1.37	1.01	1.22
Mechanical	0.595	0.819	1.11	0.764	1.14

results showed that the relation between robot-types and damage to humans and their identity was fully mediated by the human–machine distinctiveness of the different robot types.

3.2.2 Mediation Effect of Anthropomorphic Appearance on the Relation Between Robot Type and Distinctiveness

We conducted the same mediation analysis as in Study 1 on the relation between robot types as IV and human–machine distinctiveness as DV, with the anthropomorphic appearance of the different robot types as a proposed mediator. We replicate our findings in Study 1: the relation between robot-types and human–machine distinctiveness was therefore fully mediated by the anthropomorphic appearance of the different robot types.

3.2.3 Mediation Effect of Mind Attribution on the Relation Between Robot Type and Distinctiveness

We conducted a mediation analysis on the relation between robot types as IV and human–machine distinctiveness as DV, with the mind attribution of the different robot types as a proposed mediator. The model indicated non-significant indirect effects of the experience component of mind attribution, thus no mediation effect was found.

The mediation analysis that involved the agency component of mind attribution as a proposed mediator showed different results depending on which robot types were compared. In the comparisons between the mechanical versus android robot types, and between the mechanical versus humanoid robot types, the model indicated non-significant indirect effects, only significant direct effects. Furthermore, the model indicated significant indirect effects and significant direct effects in the comparison between the humanoid versus the android robot types. The relation between robot-types and human–machine distinctiveness might be partially mediated by the anthropomorphic appearance when comparing humanoid and android robots.

3.2.4 Correlation Between Mind Attribution and Damage to Humans and Their Identity

To explore whether the correlation between mind attribution and damage to humans and their identity vary differently between robot types, correlation analyses on each robot type were performed. The results showed that the correlation between agency and damage to humans and their identity for mechanical and humanoid robots were significant ($r_{\text{mechanical}} = .369$, $p = .004$; $r_{\text{humanoid}} = .313$, $p = .015$), but not for android robots ($r_{\text{android}} = .193$, $p = .140$). The correlation between experience and damage to humans and their identity for mechanical robots was significant ($r_{\text{mechanical}} = .485$, $p < .001$), but not significant for humanoid and android robots ($r_{\text{humanoid}} = .205$, $p = .116$; $r_{\text{android}} = .171$, $p = .190$).

In line with earlier research [30], we could again support the notion that an anthropomorphic appearance mediates the human/machine distinctiveness, and thus gives room to feelings of threat and damage to humans. Additionally, we did not find mediating effect of experience on participants perception of human/machine distinctiveness, while agency might partially mediate the influence on human/machine distinctiveness only for the comparison between the humanoid and android robots. Interestingly, however, is the fact that the results showed a trend of positive correlation between both experience and agency and the damage to humans and their identity.

As the correlations were not all significant, an interpretation is difficult. One could argue that especially the similarity between android robots and humans weakens a relationship between agency/experience perception with damage to humans and their identity: it is easier to identify with these robots and perceive them as closer to humans. Another explanation could be that agency and experience would be much more logical for android robots, while mechanical or humanoid robots do not need these capacities and provide a bigger threat when possessing them. To further strengthen and be able to interpret our findings of Study 2, we conducted a third study with the goal to replicate this pattern. To be sure that the current pattern is not due to the used stimuli,

new pictures of mechanical, humanoid, and android robots were introduced in Study 3.

4 Study 3

4.1 Methods

4.1.1 Participants and Design

Sixty-seven participants (57 females, 8 males, 2 unknown, $M_{age} = 19.10$ years, $SD_{age} = 1.20$, age range 17 – 23 years) completed the experiment for course credits. The experiment had a 3 (Robot type: mechanical robot vs. humanoid robot vs. android robot) within-subjects design, with damage to humans and their identity as dependent variable, and physical anthropomorphism, mind attribution, and threat to the human–machine distinctiveness as mediators. Data was acquired similar to Study 1 and Study 2.

4.1.2 Procedure and Materials

The same procedure as in Study 2 was used. However, in Study 3, new stimuli for the three robot categories (mechanical, humanoid, or android robots) were used. While mechanical robots were clearly machines (e.g., no legs), but in comparison to Study 1 and Study 2, they had arms and a head. Humanoid robots were more humanlike by having legs, arms, a torso, and a head with a face. Still, they also possess clear similarities with a machine (e.g., no hair or skin). Lastly, android robots were very high in humanlike-ness and difficult to distinguish from real humans perceptually. Only full body pictures of the robots were used in Study 3 and previously used in research [5]. Again, for each category, four different robots were used, resulting in 12 robots which had to be evaluated on agency attribution (Cronbach's $\alpha_{mechanical\ robot} = .804$; Cronbach's $\alpha_{humanoid\ robot} = .852$; Cronbach's $\alpha_{android\ robot} = .902$), experience attribution (Cronbach's $\alpha_{mechanical\ robot} = .916$; Cronbach's $\alpha_{humanoid\ robot} = .916$; Cronbach's $\alpha_{android\ robot} = .970$), anthropomorphic appearance (Cronbach's $\alpha_{mechanical\ robot} = .931$; Cronbach's $\alpha_{humanoid\ robot} = .887$; Cronbach's $\alpha_{android\ robot} = .846$), human/machine distinctiveness[3] (Cronbach's $\alpha_{mechanical\ robot} = .983$; Cronbach's $\alpha_{humanoid\ robot} = .986$; Cronbach's $\alpha_{android\ robot} = .977$), and damage to humans and their identity (Cronbach's $\alpha_{mechanical\ robot} = .902$; Cronbach's $\alpha_{humanoid\ robot} = .932$; Cronbach's $\alpha_{android\ robot} = .902$).

[3] In Study 3, for exploratory reasons three extra items were added to measure human/machine distinctiveness. However, to keep results comparable with the previous two studies, only the former introduced items were included in the analyses.

4.2 Results and Discussion

All mediation statistics (B and 99% CI) are depicted in Table 5. An overview about descriptive statistics can be found in Table 6.

4.2.1 Mediation Effect of Distinctiveness on the Relation Between Robot Types and Damage to Humans and Their Identity

We conducted the same mediation analysis as in Study 1 and Study 2: We treated robot types as IV and damage to humans and their identity as DV, and human–machine distinctiveness of the different robot types as a proposed mediator. We replicated our findings of Study 1 and Study 2: the relation between robot-types and damage to humans and their identity was therefore fully mediated by the human–machine distinctiveness of the different robot types.

4.2.2 Mediation Effect of Anthropomorphic Appearance on the Relation Between Robot Type and Distinctiveness

We conducted a mediation analysis on the relation between robot types as IV and human–machine distinctiveness as DV, with the anthropomorphic appearance of the different robot types as a proposed mediator. Differently to Study 1 and Study 2, for the comparison between the mechanical with the humanoid robot, the model indicated non-significant indirect effect. All other comparisons indicated significant indirect effects.

4.2.3 Mediation Effect of Mind Attribution on the Relation Between Robot Type and Distinctiveness

We conducted a mediation analysis on the relation between robot types as IV and human–machine distinctiveness as DV, with the mind attribution of the different robot types as a proposed mediator.

The mediation analysis that involved the experience component of mind attribution as a proposed mediator showed different results depending on which robot types were compared. Specifically, the model indicated significant indirect effects and significant direct effects in the comparison between the mechanical versus the android robot types. The relation between robot-types and human–machine distinctiveness was therefore partially mediated by difference between mechanical and android in the experience attribution. In the comparisons between the mechanical versus humanoid robot types, and the humanoid versus android robot types, the model indicated non-significant indirect effects, only significant direct effects.

Table 5 B's and 99% confidence intervals (between brackets) of the mediation analyses of Study 3, as a function of comparison (mechanical vs. humanoid; mechanical vs. android; humanoid vs. android)

Study 3 IV: Robot DV: Damage	Mediation effect	Mechanical versus android	Mechanical versus humanoid	Humanoid versus android
Mediator: Distinctiveness	Indirect effect	2.208, [1.02, 3.28]	.333, [.17, .56]	1.807, [.93, 2.75]
	Direct effect	.252, [− .90, 1.40]	.160, [.01, .31]	.161, [− .74, 1.06]
	Total effect	2.46, [1.89, 3.03]	.493, [.29, .70]	1.97, [1.47, 2.46]
Study 3 IV: Robot DV: Distinctiveness				
Mediator: Appearance	Indirect effect	3.720, [2.71, 5.42]	.188, [− .05, .45]	3.172, [2.25, 4.04]
	Direct effect	− .042, [− 1.39, 1.31]	.246, [− 01, .50]	.072, [− 1.05, 1.20]
	Total effect	3.678, [3.11, 4.24]	.434, [.22, .65]	3.244, [2.70, 3.79]
Study 3 IV: Robot DV: Distinctiveness				
Mediator: Experience	Indirect effect	.891, [.13, 1.55]	− .003, [− .05, .03]	.577, [− .05, 1.25]
	Direct effect	2.788, [1.90, 3.68]	.437, [.22, .65]	2.668, [1.86, 3.48]
	Total effect	3.678, [3.11, 4.24]	.434, [.22, .65]	3.244, [2.70, 3.79]
Study 3 IV: Robot DV: Distinctiveness				
Mediator: Agency	Indirect effect	.537, [.14, .96]	− .011, [− .08, .06]	.624, [.16, 1.13]
	Direct effect	3.141, [2.49, 3.79]	.445, [.21, .68]	2.620, [1.95, 3.29]
	Total effect	3.678, [3.11, 4.24]	.434, [.22, .65]	3.244, [2.70, 3.79]

Table 6 Descriptive statistics (means, SD) of Study 3 for all variables (N = 67)

Type	Appearance	Experience	Agency	Distinctiveness	Damage
Mean					
Android	5.57	4.00	4.63	5.32	4.57
Humanoid	1.88	2.07	2.99	2.07	2.60
Mechanical	1.43	2.10	3.44	1.64	2.11
SD					
Android	1.22	1.74	1.45	1.54	1.59
Humanoid	0.932	1.01	1.35	1.19	1.34
Mechanical	0.778	0.946	1.23	0.984	1.21

The mediation analysis that involved the agency component of mind attribution as a proposed mediator showed different results depending on which robot types were compared. Specifically, the model indicated significant indirect effects and significant direct effects in the comparison between the mechanical versus the android robot types, and the humanoid versus the android robot types. The relation between robot-types and human–machine distinctiveness was therefore partially mediated by the agency attribution of the different robot types. In the comparisons between the mechanical versus humanoid robot types the model indicated non-significant indirect effects, only significant direct effects.

4.2.4 Correlation Between Mind Attribution and Damage to Humans and Their Identity

Like in Study 2, we explored whether the correlation between mind attribution and damage to humans and their identity vary differently between robot types, correlation analyses on each robot type were performed. The results showed that the correlation between agency and damage to humans and their identity was not significant for mechanical robots ($r_{\text{mechanical}} = .128$, $p = .301$), but the correlation for the humanoid and android robots were significant ($r_{\text{humanoid}} = .259$, $p = .034$; $r_{\text{android}} = .304$, $p = .012$). The correlation between experience and damage to humans

and their identity for mechanical robots was significant ($r_{mechanical} = .252$, $p = .04$), a non-significant trend was found for humanoid robots ($r_{humanoid} = .229$, $p = .063$), and a significant correlation for android robots ($r_{android} = .309$, $p = .011$).

While we replicated earlier research on anthropomorphic appearance [30], our results on mind perception were less straightforward. Possible explanations of these findings and inconsistencies between Study 2 and Study 3 concerning mind perception will be explained in the following discussion section.

5 General Discussion

In the present study we investigated the factors that elicit negative feelings towards robots and clarified possible underlying mechanisms, including the extent to which robots look human-like and the extent to which they are attributed with a mind. More specifically, we expected that a felt decrease in distinctiveness between humans and machines would be the reason that participants feel a potential damage to their identity, and that this influence can be explained by the human-like appearance of a robot as well as the attribution of agency and experience to a robot. That is, when people attribute high levels of agency or experience to a robot, or when a robot looks very human-like, this makes the boundaries between humans and machines blurrier, leading to higher levels of perceived damage to ones' identity as human.

Importantly, we could replicate earlier research [30] which demonstrated that anthropomorphic appearance indeed mediated the relationship between type of robot, human–machine distinctiveness, and perceived damage for humans and their identity. In three studies we were able to replicate earlier findings showing that when people feel that the distinction between humans and robots are blurred intergroup distinctiveness is threatened [30], both with similar and new stimuli of mechanical, humanoid, and android robots. This further strengthens the notion that although robot familiarity can be used to reach the goal of increasing robot acceptance, "this goal should however not conflict with the need for distinctiveness that typically characterizes intergroup comparisons" [30, page 299]. Thus, when robots are designed to interact with people in their daily life, and when high acceptance is important for successful use (e.g., care taking robots for elderly), it is important to consider that human-like appearance can lead to resistance.

The findings regarding the attribution of agency and experience were less consistent than the results on anthropomorphic appearance. Based on our findings, we cannot conclude that mind attribution has comparable effects on human/machine distinctiveness as anthropomorphic appearance and mediates the relationship between type of robot, human–machine distinctiveness, and damage to humans and their identity. In Study 2, one out of six comparisons showed a mediation effect for agency when comparing humanoid and android robots, which could be the result of chance. In Study 3 more significant results were found, with agency attributions mediating the influence on human/machine distinctiveness for the comparison between the mechanical and android robots, and the humanoid and android robots. Additionally, only in study 3, experience attribution mediated the effect on participants perception of human/machine distinctiveness, but for mechanical versus android robots only. Three explanations for why we did not find a stable pattern in results for mind attribution should be mentioned: Firstly, different stimuli were used between Study 2 and Study 3, with pictures in Study 3 being more controlled (e.g., same background) and only reflecting full-body postures. Research has demonstrated that robot perception is highly dependent on the sort of stimuli, and therefore, pictures in Study 3 lead to different effects. Whether these effects are more reliable needs to be replicated in future research. Secondly, because pictures were presented, it could be that participants in our studies could not imagine that the depicted robots really possess agency or experience. Thirdly, it might be that effects of agency and experience attribution are weaker than effects of anthropomorphic appearance, and therefore, the current research had not enough power to detect this influence. Therefore, future research should use a larger sample when investigating these effects.

Interestingly, in both studies, experience and agency seemed to have some positive relation with the damage to humans and their identity, meaning that robots who were ascribed more agency and/or experience were considered more damaging to human identities. Although here an inconsistent pattern was found, the correlations fit well with research introducing an "uncanny valley of the mind" [6]: Digital agents that were perceived as autonomous artificial agents also elicited an increased feeling of eeriness in the observer. Our results suggest that it might be possible to distinguish the impact of 'mind attribution' on perceived human identity threat into two separate factors: agency and experience. Thus, attributions of agency and experience can lead to feelings of threat under certain circumstances. It can be speculated whether the ability of a robot to experience human-like emotions, such as pain or happiness, would indeed make the line between robots and humans blurrier; yet, it would also make a robot more relatable and familiar. In contrast, a robot that is able to carry out the most complex human cognitive functions without experiencing or feeling anything is not relatable to us at all. Our findings provide a first hint that these factors have a differential impact on feelings of threat to our human identity. However, definitely, further research is needed to corroborate this.

The relationship between agency, experience, and damage to humans and their identity could point to other possible mediators instead of human–machine distinctiveness. For example, the extent to which humans can relate to a robot, or feel close to a robot could alleviate the potential damage to someone's identity experienced as a result of low distinctiveness. Whether mind perception leads to a decrease in human–machine distinctiveness in moving digital avatars, or which other factors might lead to the negative feelings felt is up for future investigations.

One major limitation of the current research is that only pictures were presented, while films or interactions with robots (either real-life interactions or interactions using VR technologies, see [6]) would allow for more complexity in the design. It is therefore unclear whether seeing a robot in a picture leads to similar feelings and cognitions as seeing and interacting with an actual robot. It is likely that through real interaction, one's preliminary cognitions and feelings are updated based on experience, thus leading to more reliable and clear perceptions. Therefore, real-life interacting with robots is a promising venue for future research to explore. By using pictures, we were able to test our assumptions in a highly controlled environment, and allowed for comparisons with earlier research [30]. Nevertheless, future studies should try to implement moving materials or real interactions to further validate what factors lead to a decrease in human–machine distinctiveness and negative feelings and perceived damage to ones' identity.

Another limitation of the current study is that the design was only tested in a convenience sample of highly educated, young participants. In future research, it would be interesting to see whether the current results can also be found in a more diverse sample. As research has shown that elderly respond differently towards artificial agents [38], it might be and interesting target group: they might have less experience with seeing robots, and could therefore probably show stronger feelings of threat towards robots. In addition, it would be interesting to test how this relationship can be minimised. For example, research has shown that highlighting a shared goal can lead to less intergroup conflict [39]. Thus, making it clear that an interaction with a robot is necessary to reach a desired state might help to increase acceptance.

Our findings have important practical implications: Although a more human-like appearance and mind attribution can be beneficial for human–robot interactions [21], robots are better accepted when perceptual differences in appearance remain preserved. Thus, when designing robots for daily use, it is important users can clearly perceive non-human–robotic features in appearance and behaviour. This could for example be achieved by using a body that does not closely resemble human anatomy (such as for Cozmo or Roomba). Similarities in mind perception and experience are less threatening compared to similarities in appearance. The former does not affect the human–robot distinction as strongly as the latter. Nevertheless, robots that are perceived to have high agency and experience might evoke eeriness and unease. Therefore, while it is good that robots are able to demonstrate that they can make own decisions or have a basic experience of emotions and feelings (e.g., by verbally expressing understanding or explaining why they behave in a certain way), moderation is key if one wishes to avoid negative evaluations.

Funding This study received no funding.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Epley N, Waytz A, Cacioppo JT (2007) On seeing human: a three-factor theory of anthropomorphism. Psychol Rev 114:864–886
2. Lee KM, Jung Y, Kim J, Kim SR (2006) Are physically embodied social agents better than disembodied social agents? The effects of physical embodiment, tactile interaction, and people's loneliness in human robot interaction. Int J Hum Comput Stud 64:962–973. https://doi.org/10.1016/j.ijhcs.2006.05.002
3. Kamide H, Mae Y, Kawabe K, Shigemi S, Arai T (2012) A psychological scale for general impressions of humanoids. In: 2012 IEEE international conference on robotics and automation (ICRA), pp 4030–4037. https://doi.org/10.1080/01691864.2013.751159
4. Mori M (1970) The uncanny valley. Energy 7:33–35
5. Rosenthal-von der Pütten AM (2014) Uncannily Human. Empirical Investigation of the Uncanny Valley Phenomenon. Dissertation Thesis
6. Stein J-P, Ohler P (2017) Venturing into the uncanny valley of mind—the influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. Cognition 160:43–50. https://doi.org/10.1016/j.cognition.2016.12.010
7. Strait MK, Aguillon C, Contreras V, Garcia N (2017) The public's perception of humanlike robots: Online social commentary reflects an appearance-based uncanny valley, a general fear of a "Technology Takeover", and the unabashed sexualization of female-gendered robots. In: 2017 26th IEEE international

symposium on robot and human interactive communication (RO-MAN), Lisbon, pp 1418–1423
8. Dautenhahn K, Nehaniv CL, Walters ML, Robins B, Kose-Bagci H, Mirza NA, Blow M (2009) KASPAR–a minimally expressive humanoid robot for human–robot interaction research. Appl Bionics Biomech 6:369–397
9. Krach S, Hegel F, Wrede B, Sagerer G, Binkofski F, Kircher T (2008) Can machines think? Interaction and perspective taking with robots investigated via fMRI. PLoS ONE 3(7):e2597. https://doi.org/10.1371/journal.pone.0002597
10. Krämer NC, Kopp S, Becker-Asano C, Sommer N (2013) Smile and the world will smile with you—the effects of a virtual agent's smile on users' evaluation and behavior. Int J Hum Comput Stud 71:335–349. https://doi.org/10.1016/j.ijhcs.2012.09.006
11. Müller BCN, Brass M, Kühn S, Tsai CC, Nieuwboer W, Dijksterhuis A, van Baaren RB (2011) When Pinocchio acts like a human, a wooden hand becomes embodied. Action co-representation for non-biological agents. Neuropsychologia 49:1373–1377
12. Morewedge CK, Preston J, Wegner DM (2007) Timescale bias in the attribution of mind. J Personal Soc Psychol 93:1–11
13. Burgoon JK, Bonito JA, Bengtsson B, Cederberg C, Lundeberg M, Allspach L (2000) Interactivity in human-computer interaction: a study of credibility, understanding, and influence. Comput Hum Behav 16:553–574
14. Heider F, Simmel M (1944) An experimental study of apparent behavior. Am J Psychol 57:243–259
15. Castelli F, Happé F, Frith U, Frith C (2000) Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. Neuroimage 12:314–325
16. Iacoboni M, Lieberman MD, Knowlton BJ, Molnar-Szakacs I, Moritz M, Throop CJ, Fiske AP (2004) Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. NeuroImage 21:1167–1173
17. Kühn S, Brick TR, Müller BCN, Gallinat J (2014) Is this car looking at you? How anthropomorphism predicts fusiform face area activation when seeing cars. PLoS ONE 9:e113885
18. Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci 36:181–204
19. Friston K (2003) Learning and inference in the brain. Neural Netw 16:1325–1352. https://doi.org/10.1016/j.neunet.2003.06.005
20. Müller BCN, van Baaren RB, van Someren DH, Dijksterhuis A (2014) A present for Pinocchio: on when non-biological agents become real. Soc Cogn 32:381–396
21. Nijssen SRR, Müller BCN, van Baaren RB, Paulus M (2019) Saving the robot or the human? Robots who feel, deserve moral care. Soc Cogn 37:41–52
22. Müller BCN, Oostendorp AK, Kühn S, Brass M, Dijksterhuis A, van Baaren RB (2015) When triangles become human: action co-representation for objects. Interact Stud 16:54–67
23. Goetz J, Kiesler S, Powers A (2003) Matching robot appearance and behavior to tasks to improve human-robot cooperation. In: Proceedings of the 12th IEEE international workshop on robot and human interactive communication, pp 55–60
24. Kiesler S, Goetz J (2002) Mental models and cooperation with robotic assistants. In: Proceedings of CHI'02 on human factors in computing systems, pp 576–577
25. Gong L (2008) How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. Comput Hum Behav 24:1494–1509
26. Eyssel FA, Hegel F (2012) (S)he's got the look: gender-stereotyping of social robots. J Appl Soc Psychol 42:2213–2230
27. Bartneck C, Kanda T, Ishiguro H, Hagita N (2009) My robotic doppelganger—a critical look at the uncanny valley theory. In: Proceedings of the 18th IEEE international symposium on robot and human interactive communication, pp 269–276
28. Hanson D (2006) Exploring the aesthetic range for humanoid robots. In: Proceedings of the ICCS/CogSci-2006 long symposium: toward social mechanisms of android science, pp 16–20
29. Saygin AP, Chaminade T, Ishiguro H, Driver J, Frith CD (2012) The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. Soc Cogn Affect Neurosci 7:413–422. https://doi.org/10.1093/scan/nsr025
30. Ferrari F, Paladion MP, Jetten J (2016) Blurring human-machine distinctions: anthropomorphic appearance in social robots as a threat to human distinctiveness. Int J Soc Robot 8:287–302
31. MacDorman KF, Entezari SO (2015) Individual differences predict sensitivity to the uncanny valley. Interact Stud 16:141172. https://doi.org/10.1075/is.16.2.01mac
32. Jetten J, Spears R, Manstead AS (1996) Intergroup norms and intergroup discrimination: distinctive self-categorization and social identity effects. J Personal Soc Psychol 71:1222. https://doi.org/10.1037/0022-3514.71.6.1222
33. Jetten J, Spears R, Manstead AS (1997) Distinctiveness threat and prototypicality: combined effects on intergroup discrimination and collective self-esteem. Eur J Soc Psychol 27:635–657. https://doi.org/10.1002/(SICI)1099-0992(199711/12)27:63.0.CO;2-%23
34. Gray HM, Gray K, Wegner DM (2007) Dimensions of mind perception. Science 315(5812):619
35. Inquisit 4 [Computer software] (2017). http://www.millisecond.com
36. Montoya AK, Hayes AF (2017) Two-condition within-participant statistical mediation analysis: a path-analytic framework. Psychol Methods 22:6
37. Gray K, Wegner DM (2012) Feeling robots and human zombies: mind perception and the uncanny valley. Cognition 125:125–130
38. Müller BCN, Chen S, Nijssen SRR, Kühn S (2018) How (not) to increase elderly's tendency to anthropomorphise in serious games. PLoS ONE 13(7):e0199948
39. Dovidio JF, Love A, Schellhaas FMH, Hewstone M (2017) Reducing intergroup bias through intergroup contact: twenty years of progress and future directions. Group Process Intergroup Relat 5:606–620. https://doi.org/10.1177/1368430217712052

Dr. Barbara C. N. Müller received her PhD in 2013 from Radboud University Nijmegen. Her research focus is on human/non-human interactions, the use of robots in daily life, and how interactions with artificial agents can be improved. Currently, she works as an assistant professor at the Communication & Media department of the Behavioural Science Institute, Radboud University Nijmegen.

Xin Gao received his MSc degree in Behaviour Science from Radboud University. Currently, he works at the Marketing and Consumer Behaviour department as a PhD candidate, Wageningen University. He investigates consumers' responses to innovative products that are incongruent to their knowledge (such as bamboo textile).

Sari R. R. Nijssen having obtained her MSc degree from University College London, is currently pursuing a PhD at Radboud University Nijmegen. She works in the Department of Communication & Media, where she investigates the development of anthropomorphism in the context of human-robot interaction.

Dr. Tom G. E. Damen is an Assistant Professor at the Department of Psychology at Utrecht University. His research explores when agency and responsibility emerge, not only in fundamental contexts but also in applied settings.