Why Context Matters: The Influence of Application Domain on Preferred Degree of Anthropomorphism and Gender Attribution in Human–Robot Interaction

Eileen Roesler¹ · Lara Naendrup-Poell¹ · Dietrich Manzey¹ · Linda Onnasch²

Accepted: 23 December 2021 / Published online: 24 January 2022
© The Author(s) 2022

Abstract

The application of anthropomorphic design features is widely believed to facilitate human–robot interaction. However, the preference for robots' anthropomorphism is highly context sensitive, as different application domains induce different expectations towards robots. In this study the influence of application domain on the preferred degree of anthropomorphism is examined. Moreover, as anthropomorphic design can reinforce existing gender stereotypes of different work domains, gender associations were investigated. Therefore, participants received different context descriptions and subsequently selected and named one robot out of differently anthropomorphic robots in an online survey. The results indicate that lower degrees of anthropomorphism are preferred in the industrial domain and higher degrees of anthropomorphism in the social domain, whereas no clear preference was found in the service domain. Unexpectedly, mainly functional names were ascribed to the robots and if human names were chosen, male names were given more frequently than female names even in the social domain. The results support the assumption that the preferred degree of anthropomorphism depends on the context. Hence, the sociability of a domain might determine to what extent anthropomorphic design features are suitable. Furthermore, the results indicate that robots are overall associated more functional, than gendered (and if gendered then masculine). Therefore, the design features of robots should enhance functionalities, rather than specific gendered anthropomorphic attributes to avoid stereotypes and not further reinforce the association of masculinity and technology.

Keywords Application domain · Anthropomorphism · Gender associations · Social robots · Industrial robots · Service robots

1 Introduction

The idea of human-like machines that free mankind from labor has been addressed extensively since decades in literature, movies and research. The transfer of human-like features to non-living objects like machines that accompanies this idea is referred to as anthropomorphism and has been widely adapted in different forms and contexts in human–robot interaction (HRI) [1]. The main advantage of this multifaceted design approach is the activation of human–human interaction schemes to form appropriate expectations [2], improve coordination strategies [3] or increase empathy

[4] towards robotic interaction partners. Especially, current research in social HRI shows the constant trend that anthropomorphic design can facilitate the building of meaningful relationships to achieve a more fluent and socially situated interaction [5]. Even though the term robot seems to be intuitively bonded with an association of anthropomorphism, the question arises, whether this really is desirable in all working environments.

In work-related interactions, anthropomorphism of robots, implemented via appearance, communication or movement style and robot description [1], can lead to an underestimation of the functional character and perceived value of the robot for task fulfillment [6–8].

Moreover, facial features [9] or anthropomorphic personality descriptions [10] can reinforce existing gender stereotypes of different work domains.

In conclusion, anthropomorphism might not be universally beneficial, as it can elicit a violation of the formed

✉ Eileen Roesler
eileen.roesler@tu-berlin.de

¹ Technische Universität Berlin, Marchstr. 12, 10587 Berlin, Germany

² Humboldt-Universität zu Berlin, Rudower Chaussee 18, 12489 Berlin, Germany

expectations [11] as well as a promotion of gender stereotypes [9, 10] in specific domains.

1.1 Anthropomorphism in different contexts

The utility of anthropomorphism as a supporting design feature seems to depend on the context in which robots are used [6, 8]. Whereas positive effects emerge in social HRI, anthropomorphic features can mask the functionality of and decrease the trust in robots in task-related interactions [7, 8]. Those arbitrary effects of interaction context seem to mirror people's preferences for robot design in specific domains.

One of the most essential triggers of anthropomorphism is the visual appearance of a robot [12]. As one of the first, Goetz et al. [13] investigated, if anthropomorphism determines for what jobs robots are preferred. The results showed that participants preferred more anthropomorphic robots for jobs that required more sociability, whereas more machine-like robots were favored for jobs with less sociability. Thus, this study indicates that the appearance of the robot should match the nature of the task. Złotowski et al. [14] investigated the relationship of anthropomorphism and specific tasks more detailed for the social and service domain. The result showed that a more anthropomorphic robot that was easily distinguishable from humans, was preferred for occupations where social skills played a key role (e.g., child-minding). The machine-like robot with unfamiliar, but not threatening appearance was perceived as optimal for “dull and dirty tasks” (e.g., cleaning).

The results of the mentioned studies are a first indicator that the preferred degree of anthropomorphism varies between application domains. Nonetheless, the results are hardly generalizable since the robotic appearance was manipulated on multiple dimensions like age and gender [13] or threatening and non-threatening [14]. Furthermore, the existing research focuses more on specific tasks in the social (and partially service) domain, than on the differences between global application domains like industrial, service and social HRI. Building on earlier research [13, 14] it is assumed that robot design should match the according sociability in the respective domain. Based on the exemplary discussed research we assume that robots with different levels of anthropomorphism are preferred in different domains.

- H1:* In the industrial domain robots with a low level of anthropomorphism are preferred.
- H2:* In the service domain robots with a medium level of anthropomorphism are preferred.
- H3:* In the social domain robots with a high level of anthropomorphism are preferred.

In addition to explicit preference judgments, differences should also show up in more indirect measures. Socio-

psychological findings suggest that faster responses occur when concepts are more strongly associated [15, 16]. Therefore, we assume faster responses for the associated level of anthropomorphism in the according domain.

- H4:* In the industrial domain robots with a low level of anthropomorphism are chosen most quickly.
- H5:* In the service domain robots with a medium level of anthropomorphism are chosen most quickly.
- H6:* In the social domain robots with a high level of anthropomorphism are chosen most quickly.

Additionally, the preferred level of anthropomorphism might not be as explicitly associated in the service domain as in the industrial or social domain. Because this domain includes many diverse tasks ranging in their sociability [14], we assume longer response times in this domain compared to the others.

- H7:* The overall response time is higher in the service domain compared to the social and industrial domain.

1.2 Gender stereotypes

The industrial and social domain in particular are associated with gender stereotypical occupational roles [17]. Basfirinci et al. [18], for example, used a scenario technique, where participants had to assign names to persons only described by their occupational role. Whereas, most occupations were implicitly perceived as masculine, the job of a nurse was perceived feminine. The results illustrated that implicit occupational stereotypes still exist to date. In order to prevent further task-specific consolidation as well as transfer of stereotypes to robots, anthropomorphic design choices should be made with caution [9]. Tay et al. [17] provide empirical support for this claim. The study revealed that a robot with female features was more accepted and positively evaluated in healthcare, while a robot with male features was preferred as security guard. Additionally, a study by Ladwig and Ferstl [19] showed that humanoid robots are implicitly assigned a gender. However, the transfer of human stereotypes to robots has not been found in all cases. Rea et al. [20], for example, did not find an effect of the robot's gender on its perceived suitability for several stereotypically male or female tasks, though they note themselves that the robot's gender manipulation changing only the pronoun may not have been sufficient. In line with that, other research which gendered robots only via name and voice did not find main effects of robot gender on the acceptance of the robot for stereotypically female and male tasks [21].

Overall, research on gender stereotypes in robots reports controversial findings [22]. It is therefore necessary to conduct further systematic research.

The industrial context includes mainly technical tasks traditionally associated with men rather than women [9, 23]. In the service context, less strong stereotypes may exist, but as Ladwig and Ferstl [19] point out, robots are generally given male names more often. In contrast and as different studies involving healthcare stereotypes suggest [17, 18], the predominantly associated gender in this area is female. Hence, we assume that the perception of robots as male or female is triggered by the associated gender stereotype prevalent for jobs in the respective domain.

- H8:* In the industrial domain, male names are chosen more frequently for robots than female names.
- H9:* In the service domain, male names are chosen more frequently for robots than female names.
- H10:* In the social domain, female names are chosen more frequently for robots than male names.

As stereotypes are a sensitive issue that is associated with response biases like social desirability [15, 24], this study additionally assessed implicit associations at the end of the experiment via the implicit association test (IAT) to tap unintentional and uncontrolled responses [25]. Therefore, we investigated associations between the application fields and already female or male looking robots. The IAT is an instrument developed by Greenwald and colleagues [15] for measuring implicit attitudes. It is nowadays used in diverse scientific fields including HRI [26–28] to uncover underlying, implicit stereotypes and associations [15, 23]. Though the IAT has been criticized for poor psychometric properties [16, 29, 30], it is still accepted as a valid measure for the relative association strength between constructs [30, 31]. Based on the expected association of gender and context, it is assumed that the association strength is higher for stereotypically matched domains and robot appearances.

- H11:* A strong association between the industrial context and male robots exists.
- H12:* A strong association between the social context and female robots exists.

Overall, the study investigated the influence of the application field (industrial vs. service vs. social) on the preference for the degree of anthropomorphism and the ascribed gender of a robot, in order to allow a domain-overlapping comparison of robot preferences and existing gender stereotypes.

2 Method

2.1 Participants

Due to the pandemic-related constraints for laboratory research, the study was performed as an online study. Based on an a priori power analysis using GPower [32] with a power $(1 - \beta)$ set at 0.90 and $\alpha = 0.05$ the targeted sample size was 117. Of all 169 participants who started the study, 36 dropped out before finishing the study. This drop-out rate of 21.3% is comparable to other online studies in German-speaking regions [33]. From a total of 133 subjects participating in the study, six participants had to be excluded because of invalid scores in the IAT. Furthermore, eight subjects were excluded after a visual examination of response time outliers, resulting in a final sample of 119 participants. The sample consisted of 73 participants that were recruited via the local university participant pool and received course credit, as well as 46 participants that were recruited from the platform Prolific, receiving a small monetary compensation for participation which was calculated on the basis of the German minimum wage (3,36 € for 20 minutes). The same inclusion criteria regarding age, ranging from 18 to 50, and German nationality as well as first language were used for both recruitment approaches. Besides those criteria no restrictions were made for participation. As a result the sample represents various domains to allow more extensive insights with regard to the expectations of general users in terms of preferences and public stereotypes.

For both sampling strategies participants did not differ in terms of age ($M_{Pool} = 26.62$, $SD_{Pool} = 4.62$, $M_{Prolific} = 28.61$, $SD_{Prolific} = 7.54$, $p = 0.112$), or gender (Pool 59% female, Prolific 43% female, $p = 0.146$) nor with regard to the control variable tendency to anthropomorphize ($M_{Pool} = 43.29$, $SD_{Pool} = 11.62$, $M_{Prolific} = 42.15$, $SD_{Prolific} = 11.76$, $p = 0.608$). Taken together, participants of both recruitment strategies were on average 27.38 years old ($SD = 5.97$) and 53% of them identified themselves as female. Moreover, participants were asked about their profession and how they would classify their own professional background. Most of the participants were students (57.98%) or employees (32.77%). The sample included people with backgrounds in the industrial domain ($n = 36$), in the service domain ($n = 20$), in the social domain ($n = 36$) and in other domains ($n = 27$).

Table 1 Translated context descriptions for the industrial, service, and social domain [originally presented in German, accessible via <https://osf.io/6zq9e/> (OSF)]

Context	Description
Industrial	For the work in the industrial domain, a highly automated support will be used in the near future. The robotic assistance helps humans to assemble products. The system provides the operator with means of production, moves objects from one workstation to the next and places individual parts in the designated areas
Service	For work in the service context, a robotic assistance will soon be used. The system delivers goods to the respective destination, sorts parcels into designated areas and cleans the work surfaces. The highly automated assistance can also help staff to hand over goods to customers
Social	For work in the social environment, a highly automated assistance system will be introduced shortly. This system provides support in caring for fellow human beings on an organizational, social and emotional level. The robotic assistance can also be used for social interactions, such as sports exercises or joint leisure activities

2.2 Task and materials

2.2.1 Context descriptions

The context descriptions were generic textual representations of context-specific joint human–robot interactions (Table 1). The description of the industrial field of application included a robotic assistance that supported workers with assembling products, moving objects from one workstation to another and placing parts in designated areas. In the service context the robotic system delivered goods to a respective destination, sorted parcels into designated areas, cleaned work surfaces and supported employees in potential customer care. The social context described a robot that supported workers in the care of other people on an organizational, social and emotional level and could be used for social interactions such as sport exercises.

2.2.2 Robot stimuli

The ABOT (Anthropomorphic roBOT) Database was used to select robots with different degrees of anthropomorphism. This database contains over 250 standardized images of existing robots with differently anthropomorphic features with every robot having a score ranging from zero to 100 to indicate the degree of the robot's anthropomorphism [34]. This overall score contains four dimensions of robot appearance features that were identified with a principal component analysis: the surface features (e.g., skin, gender, hair, eyebrows),

Fig. 1 Examples of the robots used in this study with low (row 1), medium (row 2), and high (row 3) degrees of anthropomorphism

facial features (head, face, eyes, mouth), body-manipulators (e.g., legs, arms, torso) and mechanical locomotion (treads, wheels). Following the approach of previous research [13, 35, 36], three different degrees of anthropomorphism represented by the overall score were considered for the study (low, medium, high). For every anthropomorphism degree three robots were chosen to minimize carryover effects within each domain, as each context description was presented three times. Apart from differences in perceived anthropomorphism, all robots had similar color schemes, similar abilities based on their appearance and no obvious gender cues like hairstyle [37] or body proportion [38].

See Fig. 1 for examples of low, medium and high anthropomorphic robots. The scores within each category were comparable, whereas the scores between the low ($M = 9.14$, $SD = 0.56$), medium ($M = 23.06$, $SD = 0.54$) and high ($M = 49.2$, $SD = 1.82$) level of anthropomorphism differed substantially. It was a deliberate decision to not select robots with extremely high perceived anthropomorphism, because on the one hand, these often already have an assigned gender or at least gender specific cues (like long hair or wearing a dress). On the other hand, robots that are too anthropomorphic might in general be perceived negatively and generate a feeling of uncanniness [39].

2.2.3 Implicit association test

The IAT is a computer-based discrimination task, in which subjects are asked to classify individual stimuli representing concepts or attributes as quickly as possible into four different categories by pressing two possible answer keys [16]. For the four categories suitable stimuli that are easily categorizable have to be selected [16]. Typically, IAT stimuli [15] are represented by words, but images or symbols can be used as well [16]. Because gender categories with regard to robots are difficult to realize verbally, the stimuli were implemented using images of robots with typically male and female associated features. Categories in the IAT are usually represented by eight stimuli each [15]. Therefore, eight images of male and female looking robots as well as eight images of an industrial

Fig. 2 Components of the IAT stimuli with a robot rated as male, a robot rated as female, a context picture of the industrial and social domain (left to right)

and a social context were selected. The robot stimuli were mostly derived from the ABOT database while the context stimuli were extracted from free stock image databanks (see Fig. 2 for examples of the robot and context stimuli). A pre-test was conducted to find the most suitable robot stimuli. Eighteen participants (12 female) with a mean age of 34.44 ($SD = 15.35$) years, evaluated 20 stimuli with regard to the perceived gender of the robot on a scale from zero (male) to 100 (female). The mean scores for every robot were calculated and respectively the eight most male (scores between 4.9 and 30.3) and most female (scores between 66.2 and 99.4) looking robots were selected. The final stimuli can be accessed at the Open Science Framework (OSF) via <https://osf.io/6zq9e/>.

For the analysis of the IAT the improved D-Score was calculated according to Greenwald et al. [43]. This score consists of the average response time difference between the combined stages in the IAT, thus the stage where “social” and “female robot” share an answer key and “industry” and “male robot” share the other answer key as well as the stage where this pairing is reversed (social + male robot; industry + female robot) divided by the standard deviation of the respondent’s response times in both combined stages [16]. For the exact procedure see Greenwald et al. [43].

2.3 Dependent measures

2.3.1 Control measures

Though all robots had a specific score from the ABOT database, qualifying them as stimuli with low, medium or high anthropomorphism, it was still necessary to verify that the participants did perceive the differences in anthropomorphic robot design. A single item was therefore used as a manipulation check to assess the perceived anthropomorphism for each robot. The nine robots were displayed in a randomized order and participants had to indicate the human-likeness of each robot on a scale ranging from 0 (not at all human-like) to 100 (completely human-like). The scale was chosen to enable a compari-

son with the ABOT score which ranges from zero to 100, too.

Furthermore, to prevent confounding effects that influence the participants responses, the individual tendency to anthropomorphize was measured. Research has shown that the tendency to anthropomorphize non-human entities is not universal [40, 41]. To assess stable individual differences in this tendency, the individual differences in anthropomorphism questionnaire (IDAQ) by Waytz et al. [41] was used in the study.

2.3.2 Preferred degree of anthropomorphism

The main dependent variable to assess the preference for differently anthropomorphic robots was the *frequency* with which the different degrees of anthropomorphism were chosen with regard to each context. In addition to the frequencies of the chosen robots, the *response latency* (in milliseconds) of every selection was measured.

2.3.3 Gender attribution: naming frequencies

In order to examine gender associations in the application contexts, a naming technique was used that was derived from previous research [18, 19]. After the selection of a robot in a specific context, the participants were asked to give the robot a name. This open format was used in order to not impose answer options on the participants. Further, it opened up the possibility for the participants to not just give traditional male or female names but any name they could imagine, like neutral or technical ones, which is a tendency that has been observed by Keay [42] in the naming of robots for robot competitions. For the analysis, the names had to be coded into categories. For this purpose, the categories employed in earlier research [41] were modified and also applied here. The used categories are *female*, *male*, *nickname* (including names of unknown gender, popular robot names, typical animal names) and *functional* (including technical and mechanical qualities). Three raters coded the names in the different categories independent of each other in a first rating round. In a second round, the raters discussed and resolved ambiguities together. All three raters were associated with the department and two of the raters are authors of the paper (E. Roesler & L. Naendrup-Poell). The inter-rater reliability of the coded names after the first iteration was $\kappa_{\text{Fleiss}} = 0.74$. After discussing the diverging categories, in almost all cases an agreement was reached, resulting in a nearly perfect inter-rater reliability ($\kappa_{\text{Fleiss}} = 0.96$). In cases where no absolute agreement could be achieved, the category that was chosen by two third of the raters was selected.

2.3.4 Gender attribution: implicit association

In this work the automatic semantic association of the concepts “industry” and “social” with the concepts “male robot” and “female robot” was investigated. For the analysis, the so-called improved D-score was calculated. The D-score represents an index of the relative strength of association and consists of the response time differences between the expected association of congruent and incongruent category pairings [43]. It is assumed that response times are faster when two strongly associated concepts share an answer key (congruent pairing: industry and male robot or social and female robot) compared to less strongly associated concepts (incongruent pairing: industry and female robot or social and male robot) [15].

2.4 Study design and procedure

The study was conducted as an online survey using the platform *SoSci Survey*. Participants completed the study on their private computers without the presence of the experimenter.

First, participants were informed about the general terms and conditions. Afterwards, the procedure of the study was presented, and they were instructed that all the robots shown in this study are equipped with the same functional capabilities.

Subsequently, participants read one of three different context descriptions, whereby the presented order of the descriptions was randomly assigned to every subject. The descriptions of the industrial, service and social domain represented the levels of the factor “application field”. Since every participant read every domain description, the study consisted of a one factorial within-subjects design. After reading a context description, participants were asked to decide which robot they would prefer in this context based on three depicted robots. The displayed robots varied in their degree of anthropomorphism with three different levels: low, medium and high anthropomorphism (Fig. 1). After selecting a robot, subjects were asked to provide a name for the robot. This procedure was repeated in total nine times—three times for each application domain.

Thereafter the implicit association test was conducted. Participants were instructed that they had to do a categorization task and were then presented with the standardized instruction of the IAT [15, 16]. After the IAT, participants had to indicate how anthropomorphic they perceived the nine robots they had seen before as a manipulation check. Then the IDAQ [41] with fifteen items rated on an eleven-point scale from zero “not at all” to ten “very much” and several socio-demographic questions were filled in. The entire experiment lasted 15–20 min.

Fig. 3 Frequencies of robot selection (in case of anthropomorphism: low vs. medium vs. high) in percentage for the three context descriptions (industrial vs. service vs. social)

3 Results

3.1 Control measures

As intended, the low anthropomorphic robots were also rated least anthropomorphic ($M = 18.07$, $SD = 18.79$), medium anthropomorphic robots more anthropomorphic ($M = 44.17$, $SD = 22.94$) and high anthropomorphic robots most anthropomorphic ($M = 79.06$, $SD = 17.51$). A robust one-way repeated measures ANOVA and pairwise post-hoc tests based on trimmed means revealed that all differences were significant, $F_{(1.96, 139.13)} = 667.34$, $p < 0.001$. To control for individual differences in the tendency to anthropomorphize, the IDAQ was assessed and an individual sum score for every participant was computed. To examine whether the anthropomorphism tendency had an influence on the choice of anthropomorphic robots, it was analyzed whether within the three fields of application, individuals with a higher or lower IDAQ score were more likely to choose certain anthropomorphic robots. The only significant correlation was found in the social context, which, however, was negative, $r = -0.275$, $p = 0.003$. Accordingly, participants with a higher tendency to anthropomorphize tended to choose robots with a lower degree of anthropomorphism in the social domain.

3.2 Preferred degree of anthropomorphism

Overall, the preferred level of anthropomorphism varied substantially between the three contexts (Fig. 3).

Whereas a clear majority of participants preferred robots with a low level of anthropomorphism in the industrial context (58.82%) and robots with high anthropomorphism in the social context (84.03%), no clear preference emerged in the

Fig. 4 Response times of robot selection (in case of anthropomorphism: low vs. medium vs. high) in ms for the three contexts (industrial vs. service vs. social)

service context with 37.53%, 34.45% and 28.01% of participants preferring robots with a medium, low and high level of anthropomorphism, respectively. For the analysis non-parametric Friedman tests of differences among repeated measures were used to account for the relatedness of the groups as well as the ordinal scale level of the anthropomorphism degree. The analysis revealed a significant difference in the industrial context ($X^2_{(2,N=119)} = 80.26, p < 0.001$). Post-hoc Wilcoxon Signed Rank tests indicated that participants significantly preferred robots with a low level of anthropomorphism over robots with medium (32.77%, $r = 0.37$) and high (8.4%, $r = 0.71$) levels of anthropomorphism, as well as robots with a medium level of anthropomorphism over robots with a high ($r = 0.5$) level of anthropomorphism in the industrial context. The significant difference between degrees of anthropomorphism in the service domain ($X^2_{(2,N=119)} = 7.06, p = 0.029$) revealed only a preference for robots with a medium level of anthropomorphism over robots with a high ($r = 0.16$) level of anthropomorphism. In contrast to the industrial context, the significant differences in the social domain ($X^2_{(2,N=119)} = 164.62, p < 0.001$) showed that robots with a high level of anthropomorphism were favored over robots with low (7.28%, $r = 0.84$) and medium (8.63%, $r = 0.85$) levels of anthropomorphism.

Besides the selection frequencies, the response latencies of these choices (Fig. 3) were analyzed via one factorial ANOVAs separately for every context.

In the industrial domain, the analysis revealed a significant main effect of anthropomorphism ($F_{(2,135.87)} = 3.63, p = 0.029$). Post hoc tests using Bonferroni correction for multiple comparisons showed that the response time was significantly faster ($p = 0.029$), if a robot with a low level ($M = 4759$ ms, $SE = 343$ ms) of anthropomorphism was

chosen compared to a robot with a medium level of anthropomorphism ($M = 6076$ ms, $SE = 398$ ms). No significant differences were found for the choice of the robot with a high level of anthropomorphism ($M = 5041$ ms, $SE = 740$ ms) compared to both the low ($p = 1$) and medium ($p = 0.629$) anthropomorphic robot conditions. The analysis of the response latency in the service context showed no significant differences ($F_{(2,143.83)} = 0.17, p = 0.848$) in response times between the low ($M = 5418$ ms, $SE = 444$ ms), medium ($M = 5526$ ms, $SE = 381$ ms) and high ($M = 5758$ ms, $SE = 443$ ms) degree of anthropomorphism.

However, the selection latencies differed significantly ($F_{(2,120.58)} = 8.38, p < 0.001$) in the social context. Participants selected the robots with a high ($M = 3782$ ms, $SE = 246$ ms) level of anthropomorphism faster than robots with a medium ($M = 5704$ ms, $SE = 524$ ms, $p = 0.003$) and a low ($M = 5679$ ms, $SE = 630$ ms, $p = 0.016$) level of anthropomorphism. In total, the analysis revealed a significant main effect of context ($F_{(2,442.7)} = 9.44, p < 0.001$). Post hoc tests using Bonferroni correction for multiple comparisons showed that the response time in the social domain ($M = 4319$ ms, $SE = 271$ ms) was significantly faster than the response times in the industrial ($M = 5266$ ms, $SE = 253$ ms, $p = 0.007$) and service ($M = 5620$ ms, $SE = 251$ ms, $p < 0.001$) domain.

3.3 Gender associations

The naming of the selected robots (Fig. 5) showed a domain-overlapping preference for functional robot names like “industrial helper”, “liftbot” or “helpbot”. More precisely, the analysis of the industrial context revealed significant differences between the selected names ($X^2_{(3,N=119)} = 85.32, p < 0.001$). Bonferroni corrected post hoc pairwise comparisons showed that female names (2.54%) were chosen significantly less often than male names (16.67%, $p = 0.003$), nicknames (20.9%, $p < 0.001$) and functional (59.89%, $p < 0.001$) names. Furthermore, functional names were given significantly more often than male names ($p < 0.001$) and nicknames ($p < 0.001$).

The results were comparable in the service domain ($X^2_{(3,N=119)} = 52.69, p < 0.001$), as female names (5.37%) were chosen again significantly less often than male names (18.08%, $p = 0.03$), nicknames (25.42%, $p < 0.001$) and functional (51.13%, $p < 0.001$) names. Additionally, again functional names were given significantly more often than male names ($p < 0.001$) and nicknames ($p = 0.008$).

Surprisingly and against the expected stereotypes, the differences between gender associations showed a similar pattern in the social domain ($X^2_{(3,N=119)} = 19.91, p < 0.001$). Again, female names (9.04%) were chosen significantly less often than male names (27.97%, $p = 0.004$), nick-

Fig. 5 Frequencies of chosen names (female vs. male vs. nickname vs. functional) in percentage for the three contexts (industrial vs. service vs. social)

names (25.42%, $p = 0.015$) and functional names (37.57%, $p < 0.001$).

3.4 Implicit associations

On average, participants had a stronger association between the concepts industry + male robot and social + female robot than vice versa (industry + female robot, social + male robot), which is apparent from the positive improved D-Score ($M = 0.82$, $SD = 0.38$). This effect was not significantly different ($t_{(114.76)} = 0.49$, $p = 0.623$) for participants identifying themselves as female ($M = 0.84$, $SD = 0.42$) or male ($M = 0.81$, $SD = 0.32$). Additionally, the IAT scores did neither correlate with the chosen gendered name in the industrial ($r = 0.05$, $p = 0.70$) nor in the social ($r = 0.15$, $p = 0.106$) context.

4 Discussion

The aim of this study was to examine the effect of context on the preferred degree of anthropomorphism and ascribed gender in HRI. Therefore, the selection of robots with different degrees of anthropomorphism as well as the naming of those robots in three different contexts were analyzed.

The assumed differences in the preferred level of anthropomorphism between the contexts were mostly supported by the results. As expected in *H1*, participants showed a clear preference for a robot with a low level of anthropomorphism in the industrial context. This result is in line with previous research indicating that low levels of anthropomorphism are preferred for robots performing “dull and dirty tasks” [14]. Previous results already suggest that anthropomorphic robots might not be always beneficial for task performance in the industrial domain [7, 8]. Going beyond these results, our

research further suggests that the anthropomorphic design of robots does not even seem to be desired in the industrial domain. This is especially interesting as the current implementations of robots in this domain show a contrary trend. Particularly, collaborative robots (as characterized in our context description) are equipped with anthropomorphic features (e.g., Sawyer/ Baxter from HAHN Group or the Workerbot from Pi4 robotics) to be more preferred by users. This assumption is not supported by our findings. Moreover, it is questionable whether this design approach is effective as those anthropomorphic features are implemented in a task irrelevant manner (e.g., randomly moving eyes) [8].

Also, in line with *H3* and previous research [13, 14], the results were contrary in the social domain in which robots with the highest level of anthropomorphism were significantly preferred over low and medium anthropomorphic robots. Both, the result of the social and industrial domain, are supporting the matching hypothesis [13, 14] stating that robot design should match the according sociability in the respective domain.

However, the preferences were less clearly pronounced as expected in *H2* in the service domain, where only the medium anthropomorphic robots were preferred significantly over the high anthropomorphic ones. One possible reason for the ambiguous preferences in the service domain might be that the written vignette allowed for a wide range of interpretation.

The service domain includes many diverse tasks ranging in their sociability [14]. Participants might have developed different ideas about the robots’ scope of duty and the required sociability.

The preference for anthropomorphism in different domains therefore might have been moderated through the expected sociability of the task. Future research is thus needed to further detail the preferred degree of anthropomorphism by taking the domain and the specific task sociability into account.

In addition, the association of anthropomorphism with the sociability of domains is supported by the analysis of the response time data. It was assumed that faster responses occur when concepts are more strongly associated [15, 16]. Whereas participants reacted fastest to the low anthropomorphic robot in the industrial domain (*H4*), the fastest response times in the social domain were found for the robot with a high degree of anthropomorphism (*H6*). Again, our expectation in the service domain was not met (*H5*), as no clear differences in response times occurred in this domain.

Moreover, the overall response times were faster in the social domain (with predominantly chosen robots with a high degree of anthropomorphism) compared to the service and industrial domain. This result only partially supports *H7*. We assumed that the choice in both, the social and industrial domain, would be faster than in the service domain

as the association of a degree of anthropomorphism was expected to be higher in those domains. For the social domain there indeed seems to be a strong association between the social field of application and highly anthropomorphic robots, which is in line with fast response times revealed in different studies [15, 16]. In the industrial domain the association with low anthropomorphism might not be as clearly present as expected, which was additionally supported by the higher variability of the choice frequency in the industrial domain compared to the social domain.

In summary, anthropomorphism is associated and preferred in the social domain, but not in the industrial one. This result empathizes that the transfer of features, which facilitate social HRI [3] might not be always beneficial in other contexts.

However, it is important to keep in mind that the performed online study limited the possible ways of perceiving robots to a visual depiction only and thus important interactional factors of HRI associated with physical embodiment like movements, sounds or communication could not be experienced [1]. As commonly emphasized in HRI research [20, 26] there is a strong need for real interaction studies and actual field studies as the generalizability of online studies is not always clear cut due to the low external validity. Studies with embodied robots therefore represent a necessary requirement to further investigate people's preferences when they actually interact with a robot and whether environmental aspects like the social and organizational context influence this preference. Furthermore, it is necessary to systematically investigate the influence of other dimensions of anthropomorphism. Anthropomorphism in this study was implemented via the appearance of the robot but anthropomorphism can go beyond appearance by including movements, communication or the context [1].

Additionally, the negative correlation of the individual tendency to anthropomorphize [41] and the selection of robots in the social domain is noteworthy. It could be presumed that individuals with a higher tendency to anthropomorphize robots perceived robots in general as more anthropomorphic and therefore even the lower anthropomorphic robots were perceived as suitable in a social context. This assumption needs to be addressed in future research, to supplement general design recommendations with knowledge about interindividual preferences [40, 41].

Nonetheless, the results enable a first domain overlapping comparison. Whereas earlier research investigated the preference for robot anthropomorphism on specific tasks and primarily in the social and service domain [13, 14], the current study examined the preferred level of anthropomorphism under situational circumstances of mentioned domains and additionally the industrial domain. Hereby, it is important to state that the context descriptions clearly presented all robots with similar capabilities and the same interactional

aim of collaboration. This standardized approach with regard to robot capabilities and task relevance further supports the meaning of application domains as a central influencing factor for HRI. Overall, the results strengthen the assumption that the preferred morphology of the robot is depending on the expected social functions needed in the different domains [13]. Anthropomorphic robots are preferred for the social domain, in which tasks are more associated with human communication behavior and low anthropomorphic robots are favored in the industrial domain, in which tasks are more associated with physical demands.

Surprisingly and contrary to previous research [9, 17, 18, 23], the analysis of gender associations revealed that a clear majority of participants always chose a functional name for the robots in the different contexts. Furthermore, and in line with *H8* and *H9*, participants choosing a gendered name preferred male names over female names, in the service and industrial context. Unexpectedly and against *H10* male names were also chosen more frequently than female ones in the social domain. This latter result indicates a generally stronger association of technology with masculinity than with femininity [22]. Future research needs to investigate this assumption of a possible robot-masculinity bias by comparing the ascribed gender of robots with and without explicit gender cues. Noteworthy, however, gendered human names (either male or female) for robots were more often considered in the social (37%), than in the service (23.45%) and industrial (19.12%) domain. This is in line with the clearly preferred anthropomorphic robot design in the social domain.

Nevertheless, the association that is prevalent in all contexts seems to be technical or functional rather than gendered. This is an interesting finding as it challenges the existing trend to gender robots [44] because it is supposed to be socio-economically profitable [22]. Gendering robots according to human stereotypes is re-enforcing those stereotypes and thus, an ethically questionable approach. The IAT results in this work indicate that the assumed (*H11* and *H12*) stereotypical gender-occupation associations already exist in HRI, which is alarming. However, it is important to state that the participants of this study were mainly young adults with different professional backgrounds. On the one hand, this enables insights on a broader public preference for anthropomorphism and stereotypes in HRI. On the other hand, further research is needed to investigate if those results hold true for professionals in the respective domain.

The pressing question of whether such a stereotypical fit between gender and task is even necessary was addressed in a study by Dufour and Nihan [19]. They revealed that stereotypical judgment and perception of robots could be mostly diminished by providing technical features and skills of the robot. Moreover, Bryant et al. [45] showed that a gender-neutral robot brings no disadvantages in comparison to a gendered one. Those insights along with the results of this

work, indicating that the prevalent association irrespective of the application field is a functional one, support the opportunity for designing robots as functional and ungendered entities. Therefore, framing robots functionally and emphasizing their technical features by e.g., not giving gendered human names should be the favored alternative since it is not ethically concerning and perpetuating stereotypes.

5 Conclusions

Overall, the results of the study are in line with previous research about anthropomorphism preferences in different occupational fields, by showing that anthropomorphism is desired in social but not industrial domains. Our study therefore strengthens the body of research, which points out that anthropomorphic features are not domain-overlapping desired by the general public.

Further the results suggest that mainly a functional association of robots in the public perception prevails. However, if a gender association occurs, it is, regardless of the context, predominantly an association of masculinity with robots, and not as expected a replication of context specific gender stereotypes. Nonetheless, the predominantly functional naming of the robots in our study without explicit gender cues shows that ungendered morphology can lead to mainly functional associations.

Robot design and framing focusing on functionalities, rather than specific gendered anthropomorphic attributes can therefore be consciously used to avoid stereotypes. Both aspects, the possibility to encourage less biased HRI and the not generalizable requirement of anthropomorphic features can facilitate design approaches in HRI.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The preregistration, all stimuli and collected data can be accessed via <https://osf.io/6zq9e/>, to ensure transparent and reproducible research.

Declarations

Conflict of interest The authors have no financial or proprietary interests in any material discussed in this article.

Ethical approval All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The complete study design was approved by the local ethic committee (Tracking number: ER_01_20200604).

Informed consent Informed consent was obtained from all individual participants for whom data is included in this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Onnasch L, Roesler E (2020) A taxonomy to structure and analyze human–robot interaction. *Int J Soc Psychol*. <https://doi.org/10.1007/s12369-020-00666-5>
2. Haring KS, Watanabe K, Mougenot C (2013) The influence of robot appearance on assessment. In: 2013 8th ACM/IEEE International conference on human–robot interaction (HRI). IEEE, pp 131–132. <https://doi.org/10.1109/HRI.2013.6483536>
3. Staudte M, Crocker MW (2011) Investigating joint attention mechanisms through spoken human–robot interaction. *Cognition* 120(2):268–291. <https://doi.org/10.1016/j.cognition.2011.05.005>
4. Nijssen SR, Müller BC, Baaren RBV, Paulus M (2019) Saving the robot or the human? Robots who feel deserve moral care. *Soc Cogn* 37(1):41–56. <https://doi.org/10.1521/soco.2019.37.1.41>
5. Duffy BR (2003) Anthropomorphism and the social robot. *Robot Auton Syst* 42(3–4):177–190. [https://doi.org/10.1016/S0921-8890\(02\)00374-3](https://doi.org/10.1016/S0921-8890(02)00374-3)
6. Darling K (2015) Who's Johnny? Anthropomorphic framing in human–robot interaction, integration, and policy. In: Lin P, Jenkins R, Abney K (eds) *Robot Ethics 2.0: from autonomous cars to artificial intelligence*. Oxford University Press, Oxford, pp 173–192. <https://doi.org/10.2139/ssrn.2588669>
7. Onnasch L, Roesler E (2019) Anthropomorphizing robots: the effect of framing in human–robot collaboration. In: *Proceedings of the human factors and ergonomics Society annual meeting*, vol 63, issue no 1. SAGE Publications, Los Angeles, pp 1311–1315. <https://doi.org/10.1177/10711813196312091>
8. Roesler E, Maier JI, Onnasch L (2020) The effect of anthropomorphism and failure comprehensibility on human–robot trust. In: *Proceedings of the human factors and ergonomics society annual meeting*
9. Eyssel F, Hegel F (2012) (S)he's got the look: Gender stereotyping of robots. *J Appl Soc Psychol* 42(9):2213–2230. <https://doi.org/10.1111/j.1559-1816.2012.00937.x>
10. Kraus M, Kraus J, Baumann M, Minker W (2018) Effects of gender stereotypes on trust and likability in spoken human–robot interaction. In: *Proceedings of the eleventh international conference on language resources and evaluation*
11. Fink J (2012) Anthropomorphism and human likeness in the design of robots and human–robot interaction. In: *International conference on social robotics*. Springer, Berlin, pp 199–208. https://doi.org/10.1007/978-3-642-34103-8_20
12. Hegel F (2012) Effects of a robot's aesthetic design on the attribution of social capabilities. In: *The 21st IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, pp 469–475. <https://doi.org/10.1109/ROMAN.2012.6343796>

13. Goetz J, Kiesler S, Powers A (2003) Matching robot appearance and behavior to tasks to improve human-robot cooperation. In: The 12th IEEE international workshop on robot and human interactive communication (RO-MAN). IEEE, pp 55–60. <https://doi.org/10.1109/ROMAN.2003.1251796>
14. Złotowski J, Khalil A, Abdallah S (2019) One robot doesn't fit all: aligning social robot appearance and job suitability from a Middle Eastern perspective. *AI & SOC* 1–16. <https://doi.org/10.1007/s00146-019-00895-x>
15. Greenwald AG, McGhee DE, Schwartz JLK (1998) Measuring individual differences in implicit cognition: the implicit association test. *J Pers Soc Psychol* 74(6):1464–1480
16. Wittenbrink B, Schwarz N (2007) *Implicit measures of attitudes*. The Guilford Press
17. Tay B, Jung Y, Park T (2014) When stereotypes meet robots: the double-edge sword of robot gender and personality in human–robot interaction. *Comput Hum Behav* 38:75–84. <https://doi.org/10.1016/j.chb.2014.05.014>
18. Basfirinci C, Uk ZC, Karaoglu S, Onbas K (2019) Implicit occupational gender stereotypes: a research among Turkish university students. *Gend Manag Int J* 34(2):157–184. <https://doi.org/10.1108/GM-07-2018-0084>
19. Ladwig RC, Ferstl EC (2018) What's in a name? An online survey on gender stereotyping of humanoid social robots. In: *Proceedings of the 4th conference on gender & IT*, pp 67–69. <https://doi.org/10.1145/3196839.3196851>
20. Rea DJ, Wang Y, Young JE (2015) Check your stereotypes at the door: an analysis of gender typecasts in social human–robot interaction. In: *International conference on social robotics*. Springer, Cham, pp 554–563. https://doi.org/10.1007/978-3-319-25554-5_55
21. Kuchenbrandt D, Häring M, Eichberg J, Eyssel F, André E (2014) Keep an eye on the task! How gender typicality of tasks influence human–robot interactions. *Int J Soc Robotics* 6(3):417–427
22. Dufour F, Ehrwein Nihan C (2016) Do robots need to be stereotyped? Technical characteristics as a moderator of gender stereotyping. *Soc Sci* 5(3):27. <https://doi.org/10.3390/socsci5030027>
23. White MJ, White GB (2006) Implicit and explicit occupational gender stereotypes. *Sex Roles* 55(3–4):259–266. <https://doi.org/10.1007/s11199-006-9078-z>
24. Devine PG (2001) Implicit prejudice and stereotyping: How automatic are they? Introduction to the special section. *J Pers Soc Psychol* 81(5):757–759. <https://doi.org/10.1037/0022-3514.81.5.757>
25. De Houwer J, Moors A (2007) How to define and examine implicit processes. In: Wittenbrink B, Schwarz N (eds) *Implicit measures of attitudes*. The Guilford Press, pp 179–194
26. de Graaf MM, Allouch SB, Lutfi S (2016) What are people's associations of domestic robots?: comparing implicit and explicit measures. In: *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, pp 1077–1083. <https://doi.org/10.1109/ROMAN.2016.7745242>
27. MacDorman KF, Vasudevan SK, Ho CC (2009) Does Japan really have a robot mania? Comparing attitudes by implicit and explicit measures. *AI & Soc* 23(4):485–510. <https://doi.org/10.1007/s00146-008-0181-2>
28. Sanders TL, Schafer KE, Volante W, Reardon A, Hancock PA (2016) Implicit attitudes toward robots. In: *Proceedings of the human factors and ergonomics society annual meeting*, vol 60, issue no 1, pp 1746–1749. <https://doi.org/10.1177/1541931213601400>
29. Rothermund K, Wentura D (2004) Underlying processes in the Implicit Association Test: dissociating salience from associations. *J Exp Psychol: Gen* 133(2):139–165. <https://doi.org/10.1037/0096-445.133.2.139>
30. Schimmack U (2019) The implicit association test: a method in search of a construct. *Perspect on Psychol Sci*. <https://doi.org/10.1177/1745691619863798>
31. Kurdi B, Ratliff KA, Cunningham WA (2020) Can the Implicit Association Test serve as a valid measure of automatic cognition? A response to Schimmack (2020). *Perspect Psychol Sci*. <https://doi.org/10.1177/1745691620904080>
32. Faul F, Erdfelder E, Lang AG, Buchner A (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 39:175–191. <https://doi.org/10.3758/BF03193146>
33. Frick A, Bächtiger MT, Reips UD (2001) Financial incentives, personal information and drop-out rate in online studies. *Dimens Internet Sci* 209–219
34. Phillips E, Zhao X, Ullman D, Malle BF (2018) What is human-like? Decomposing robots' human-like appearance using the anthropomorphic roBOT (ABOT) Database. In: *Proceedings of the 2018 ACM/IEEE international conference on human–robot interaction*. IEEE, pp 105–113. <https://doi.org/10.1145/3171221.3171268>
35. Hegel F, Krach S, Kircher T, Wrede B, Sagerer G (2008) Understanding social robots: a user study on anthropomorphism. In: *2008 17th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, pp 574–579. <https://doi.org/10.1109/ROMAN.2008.4600728>
36. Prakash A, Rogers WA (2013) Younger and older adults' attitudes toward robot faces: effects of task and humanoid appearance. In: *Proceedings of the human factors and ergonomics society annual meeting*, vol 57, issue no 1, pp 114–118. <https://doi.org/10.1177/1541931213571027>
37. Hegel F, Eyssel F, Wrede B (2010) The social robot 'flobi': key concepts of industrial design. In: *19th International symposium in robot and human interactive communication*. IEEE, pp 107–112
38. Bernotat J, Eyssel F, Sachse J (2019) The (fe) male robot: how robot body shape impacts first impressions and trust towards robots. *Int J Soc Robotics* 1–13
39. Mori M, MacDorman KF, Kageki N (2012) The uncanny valley [from the field]. *IEEE Robotics Autom Mag* 19(2):98–100
40. Epley N, Waytz A, Cacioppo JT (2007) On seeing human: a three-factor theory of anthropomorphism. *Psychol Rev* 114(4):864. <https://doi.org/10.1037/0033-295X.114.4.864>
41. Waytz A, Cacioppo JT, Epley N (2012) Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspect Psychol Sci* 5(3):219–232. <https://doi.org/10.1177/1745691610369336>
42. Keay A (2011) Emergent phenomena of robot competitions: robot identity construction and naming. In: *Proceedings of the workshop on advanced robotics and its social impacts (ARSO)*. IEEE, pp 12–15. <https://doi.org/10.1109/ARSO.2011.6301972>

43. Greenwald AG, Nosek BA, Banaji MR (2003) Understanding and using the implicit association test: I. An improved scoring algorithm. *J Pers Soc Psychol* 85(2):197. <https://doi.org/10.1037/0022-3514.85.2.197>
44. Robertson J (2010) Gendering humanoid robots: Robosexism in Japan. *Body Soc* 16(2):1–36. <https://doi.org/10.1177/1357034X10364767>
45. Bryant DA, Borenstein J, Howard A (2020) Why should we gender?: the effect of robot gendering and occupational stereotypes on human trust and perceived competency. In: Belpaeme T, Young J, Gunes, H, Riek L (eds., *Proceedings of the 2020 ACM/IEEE international conference on human robot interaction (HRI)*. IEEE, pp 13–21. <https://doi.org/10.1145/3319502.3374778>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.