Form follows Function: Challenging the Effectiveness and Desirability of Anthropomorphism in Human-Robot Interaction

vorgelegt von
Eileen Roesler
ORCID: 0000-0002-7243-4882

an der Fakultät V - Verkehrs- und Maschinensysteme
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktorin der Naturwissenschaften
- Dr. rer. nat. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Klaus Gramann
Gutachterin: Prof. Dr. Linda Onnasch
Gutachterin: Prof. Dr. Friederike Eyssel

Tag der wissenschaftlichen Aussprache: 27.01.2023

Berlin 2023

Abstract

Robots and humans are encountering each other in a plethora of professional and everyday situations. With increasing direct collaboration of both entities, new design challenges arise to enable a pleasant and smooth interaction. A popular approach to achieve this is the application of anthropomorphic design features to robots. The basis for the effectiveness of this design feature is the activation of well-known interaction schemes and expectations from human-human interaction. This is assumed to increase the intuitiveness, and consequently facilitate the interaction. However, current research apart from the social domain shows that triggering human-like expectations entails not only advantages. So far, though, it is still largely unclear under which circumstances anthropomorphism is beneficial, ineffective or even detrimental. Therefore, the objective of this thesis was to gain insights into the effectiveness of anthropomorphism. Importantly, this thesis wanted to broaden the view on anthropomorphism beyond the often focused social human-robot interaction to more task-related settings. In this effort, not only the question of effectiveness but also that of desirability should be answered. Moreover, as anthropomorphic design triggers reactions comparable to those towards other humans, social issues like gender categorization play an important role. In particular, the different application domains in which robots are implemented in might foster the transfer of occupational stereotypes. This thesis aimed to challenge the approach of implementing anthropomorphism to robots in regard to effectiveness, desirability, and the social issue of gendering robots.

To answer the question of effectiveness, a meta-analysis was conducted as the first study. This analysis established a theoretical framework of the effectiveness of anthropomorphism with possible moderators and different outcome variables. This framework was quantitatively investigated. The results of this analysis provided supporting evidence that anthropomorphism depends on various moderators. Most prominently, only the social field of application consistently profited from anthropomorphism.

Based on this insight, the second study investigated whether the effect of application domain also influences people's preferences for anthropomorphic robot design. Moreover, as application domains are linked to existing occupational stereotypes, the possible transfer of these stereotypes to robots was examined. The results were consistent with the findings of the meta-analysis, as high anthropomorphic robots were preferred in the social domain only. For the industrial domain, a clear preference for low anthropomorphic robots occurred, and for the service domain the results were mixed. Interestingly, the results showed that robots are overall connected with more functional, than gendered associations. However, if they were perceived gendered, a male-robot bias occurred independently of the domain.

To rule out the possibility that this result was mainly related to our methodology rather than an existing male-robot bias, the third study was conducted. At this point, it was unknown whether this bias was induced via a male appearance of higher anthropomorphic robots even if they did not incorporate obvious gender cues. Moreover, the native language of participants might have been an influencing factors, as in German the word robot is grammatically masculine gendered. Robots with a higher degree of anthropomorphism were perceived as significantly more male than neutral or even female. The results indicated no differences in the ascription of gender to robots between grammatically gendered (i.e., German) and gender-neutral (i.e., English) languages. However, the second experiment of this study revealed that masculine grammatical gender tends to reinforce a male ascription whereas neuter grammatical gender can lead to a reduction of the male-robot bias for gender-neutral robots. As no difference between English and German participants was found and the second study used a neuter grammatical gender for describing robots, the results illustrated that the male-robot bias found in the second study was mainly associated with the appearance of most anthropomorphic robots.

For this reason, the fourth study investigated the preferred degree of anthropomorphism and possible occupational gender stereotypes with clearly gender-neutral robots. In addition, the possible influence of task sociability generally associated with the social domain and not with the industrial domain was investigated as extension of the second study. The results showed an overall preference for low anthropomorphic robots in the industrial domain for tasks requiring both low and high sociability. In comparison, task sociability mattered for the social domain, where low anthropomorphic robots were preferred for tasks requiring low task sociability. Again, the results indicate a more functional than gendered association of robots. However, in the industrial domain a male-robot bias was found, whereas no such bias was revealed in the social domain.

From a practical point of view, the results of this thesis question the current trend of implementing anthropomorphism to robots in all domains. Moreover, even gender-neutral robots might lead to a transfer of gender stereotypes, especially if referred in a grammatically masculine gendered manner and in the industrial domain. Since this domain does not clearly benefit from decorative anthropomorphism in terms of effectiveness and desirability, this design approach should be refrained. From a theoretical point of view, the studies incorporated in this thesis extend the current body of research by a systematic overview of the effectiveness of anthropomorphism and preference in different domains. In addition, the presented research opens up multiple avenues for future research, in particular by investigating the considerable research gaps revealed in the meta-analysis and the still under researched effects of gender-neutral robot design.

Zusammenfassung

Roboter und Menschen begegnen sich in einer Fülle von beruflichen und alltäglichen Situationen. Mit zunehmender direkter Zusammenarbeit beider Entitäten entstehen neue Designherausforderungen, um eine angenehme und reibungslose Interaktion zu ermöglichen. Ein beliebter Ansatz, um dies zu erreichen, ist die Implementierung von anthropomorphen Designmerkmalen bei Robotern. Grundlage für die Wirksamkeit dieser Gestaltungsmerkmale ist die Aktivierung bekannter Interaktionsschemata und Erwartungen aus der Mensch-Mensch-Interaktion. Es wird angenommen, dass dies die Intuitivität erhöht und somit die Interaktion erleichtert. Aktuelle Forschung außerhalb des sozialen Bereichs zeigt jedoch, dass das Auslösen von menschenähnlichen Erwartungen nicht nur Vorteile mit sich bringt. Bislang ist jedoch noch weitgehend unklar, unter welchen Umständen Anthropomorphismus vorteilhaft, ineffektiv oder sogar schädlich sein kann. Ziel dieser Arbeit war es daher, differenzierte Erkenntnisse über die Wirksamkeit von Anthropomorphismus zu gewinnen. Ein wichtiger Aspekt dieser Arbeit war es, die Betrachtung des Anthropomorphismus über die häufig im Mittelpunkt stehende soziale Mensch-Roboter-Interaktion hinaus auf aufgabenbezogene Situationen auszuweiten. Dabei sollte nicht nur die Frage nach der Wirksamkeit, sondern auch nach der Erwünschtheit beantwortet werden. Da anthropomorphes Design Reaktionen auslöst, die mit denen gegenüber anderen Menschen vergleichbar sind, spielen zudem soziale Fragen wie die Geschlechterzuordnung eine wichtige Rolle. Insbesondere die unterschiedlichen Anwendungsbereiche, in denen Roboter eingesetzt werden, können die Übertragung von Berufsstereotypen begünstigen. Ziel dieser Arbeit war es daher, die Implementierung von Anthropomorphismus bei Robotern hinsichtlich der Effektivität, der Erwünschtheit und der sozialen Frage der Geschlechtszuschreibung zu Robotern zu hinterfragen.

Zur Beantwortung der Frage nach der Effektivität wurde als erste Studie eine Metaanalyse durchgeführt. In dieser Analyse wurde ein theoretisches Modell für die Effektivität von Anthropomorphismus mit möglichen Moderatoren und verschie-

denen Ergebnisvariablen aufgestellt. Dieses Modell wurde quantitativ untersucht. Die Ergebnisse dieser Analyse lieferten Belege dafür, dass Anthropomorphismus von verschiedenen Moderatoren abhängt. Am auffälligsten war, dass nur der soziale Anwendungsbereich durchgängig von Anthropomorphismus profitierte.

Basierend auf dieser Erkenntnis wurde in der zweiten Studie untersucht, ob der Anwendungsbereich auch die Präferenzen der Menschen für anthropomorphes Roboterdesign beeinflusst. Da Anwendungsdomänen mit bestehenden Berufsstereotypen verknüpft sind, wurde außerdem die mögliche Übertragung dieser Stereotypen auf Roboter untersucht. Die Ergebnisse stimmen mit den Erkenntnissen der Metaanalyse überein, da anthropomorphe Roboter nur im sozialen Bereich klar bevorzugt wurden. Im industriellen Bereich wurden eindeutig Roboter mit geringem Anthropomorphismus bevorzugt, und im Servicebereich waren die Ergebnisse gemischt. Interessanterweise zeigten die Ergebnisse, dass mit Robotern insgesamt mehr funktionale als geschlechtsspezifische Assoziationen verbunden werden. Wurden sie jedoch als geschlechtsspezifisch wahrgenommen, kam es unabhängig vom Bereich zu einem Male-Robot-Bias.

Um auszuschließen, dass dieses Ergebnis hauptsächlich auf die angewandte Methodik und nicht auf einen generellen Male-Robot-Bias zurückzuführen ist, wurde die dritte Studie durchgeführt. Zu diesem Zeitpunkt war nicht bekannt, ob die Befunde durch ein männlich wahrgenommenes Aussehen höher anthropomorpher Roboter hervorgerufen wurde, auch wenn diese keine offensichtlichen Hinweise auf ein Geschlecht enthielten. Außerdem könnte die Muttersprache der Versuchspersonen ein Einflussfaktor gewesen sein, da das Wort Roboter im Deutschen grammatikalisch männlich ist. Roboter mit einem höheren Grad an Anthropomorphismus wurden signifikant mehr als männlich wahrgenommen als neutral oder sogar weiblich. Zudem deuten die Ergebnisse darauf hin, dass es zwischen grammatikalisch geschlechtsspezifischen (Deutsch) und geschlechtsneutralen (Englisch) Sprachen keine generellen Unterschiede bei der Zuschreibung von Geschlecht zu Robotern gibt. Das zweite Experiment dieser Studie zeigte jedoch, dass die Verwendung eines männlichen grammatikalischen Geschlechts tendenziell eine männliche Zuschreibung verstärkt, während das sächliche grammatikalische Geschlecht bei geschlechtsneutralen Robotern zu einer Verringerung des Male-Robot-Bias führen kann. Da kein Unterschied zwischen englischen und deutschen Versuchspersonen gefunden wurde und die zweite Studie ein sächliches Geschlecht für die Beschreibung von Robotern verwendete, zeigten die Ergebnisse, dass der in der zweiten Studie gefundene Male-Robot-Bias hauptsächlich mit dem Aussehen der meisten anthropomorphen Roboter zusammenhängt.

Aus diesem Grund untersuchte die vierte Studie den bevorzugten Grad des Anthropomorphismus und die beruflichen Geschlechterstereotypen mit eindeutig geschlechtsneutralen Robotern. Zusätzlich wurde der mögliche Einfluss der Soziabilität der Aufgabe, die im Allgemeinen mit dem sozialen Bereich und nicht mit dem industriellen Bereich assoziiert wird, als Erweiterung der zweiten Studie untersucht. Die Ergebnisse zeigten eine allgemeine Präferenz für wenig anthropomorphe Roboter in der industriellen Domäne für Aufgaben, die sowohl geringe als auch hohe Soziabilität erfordern. Im Vergleich dazu spielte die Soziabilität der Aufgabe im sozialen Bereich eine entscheidende Rolle. Hier wurden Roboter mit geringem Anthropomorphismus für Aufgaben bevorzugt, die eine geringe Soziabilität der Aufgabe erforderten. Erneut deuten die Ergebnisse auf eine eher funktionale als geschlechtsspezifische Zuordnung von Robotern hin. Im industriellen Bereich wurde jedoch Male-Robot-Bias gefunden, während dies im sozialen Bereich nicht der Fall war.

Aus praktischer Sicht stellen die Ergebnisse dieser Arbeit den derzeitigen Trend infrage, Robotern in allen Anwendungsdomänen mit anthropomorphen Merkmalen auszustatten. Darüber hinaus könnten selbst geschlechtsneutrale Roboter zu einer Übertragung von Geschlechterstereotypen führen, insbesondere wenn sie in einer grammatikalisch männlichen Weise bezeichnet und in der industriellen Domäne implementiert werden. Da diese Domäne nicht eindeutig von dekorativem Anthropomorphismus in Bezug auf Effektivität und Einstellungen profitiert, sollte von diesem Designansatz Abstand genommen werden. Aus theoretischer Sicht erweitern die in dieser Arbeit einbezogenen Studien den derzeitigen Forschungsstand um einen systematischen Überblick über die Wirksamkeit von Anthropomorphismus und Präferenz für dieses Design in verschiedenen Anwendungsdomänen. Darüber hinaus eröffnet die vorgestellte Arbeit zahlreiche Möglichkeiten für zukünftige Forschung, insbesondere durch die Untersuchung der erheblichen Forschungslücken, die in der Meta-Analyse aufgedeckt wurden, und den noch zu wenig erforschten Auswirkungen von geschlechtsneutralem Roboterdesign.

Abbreviations

HHI	Human-Human Interaction
HRI	Human-Robot Interaction

Contents

1 General Introduction

"We look at a person and immediately a certain impression of ... [the] character forms itself in us. A glance, a few spoken words are sufficient to tell us a story about a highly complex matter."

– Solomon Asch (1946, p. 258) –

Researchers, as well as the general public, have been passionate about making interferences of intentions and capabilities of others based on their morphology for quite a while (Asch, 1946; Galton, 1878). Whereas this mechanism seems to be an intuitive way to handle the uncertainty of human-human interaction (HHI), its validity is at least questionable (Zebrowitz & Montepare, 2008). However, even though the accuracy of our impressions might not be particularly high, humans can't help but form instant impressions of their interaction counterparts. Moreover, these impressions are often shared between humans. This can lead to a considerable agreement of impressions, which further guide behavior towards the interaction partners (Zebrowitz & Montepare, 2008). For example, the impression of an attractive appearance is associated with social and intellectual competence, which can lead to more favorable behavior towards the respective person (Eagly et al., 1991). Even though appearance is probably the most salient feature of a human's morphology, the initial quotation of Asch (1946) already illustrates that other aspects like communication or movement also play a role in the formation of impressions. In addition, the context in which humans are situated during the interaction matters for the impression and subsequent behavior (Hehman et al., 2019). Thus, we can state that humans, based on multiple aspects of the morphology of others, generate information about their counterpart and the social interaction they attend.

This process is not limited to HHI but also takes place in human-technology interaction, as humans perceive and react toward technologies as social actors (Nass

& Moon, 2000). Technologies that particularly evoke those social responses due to their embodiment are robots (Breazeal, 2003; Złotowski et al., 2014). Moreover, social responses are purposely triggered by designers via anthropomorphic design features of robots (Duffy, 2003; Fischer, 2022). This is done, as anthropomorphic design is assumed to facilitate the human-robot interaction (HRI) due to two major reasons. First, as robots are integrated into the natural environment of humans, anthropomorphic design serves as a functional embodiment, optimized for this environment (Fitzpatrick et al., 2016; Fong et al., 2003). Second, and more important from a psychological perspective, anthropomorphism is assumed to activate HHI scripts, which enable intuitive interactions with novel technologies like robots (Fischer, 2022). However, recent research challenges the generalizability of this assumption. While there is converging evidence that anthropomorphism can improve the HRI in the social domain (e.g., robots as interaction partners in elderly care) (Breazeal, 2003; Fong et al., 2003), it seems to undermine the perceived functionality and reliability of robots in industrial settings (Onnasch & Hildebrandt, 2021; Roesler et al., 2020). The current research landscape therefore already allows for making speculations that the application domain and other factors like task-relatedness might influence the effectiveness of anthropomorphism as a design feature. Nonetheless, there is a lack of integration of individual study results that would allow for generalized knowledge about factors that influence the effectiveness of anthropomorphism. This is a particularly urgent problem as there is still an increasing tendency across various contexts and task scenarios to implement anthropomorphic features to robots (Millo et al., 2021).

Perhaps one of the most pertinent questions that emerged in HRI research is under which circumstances anthropomorphic morphology is beneficial and desirable for HRI. This thesis aims to address this question by systematically: (1) identifying possible influencing factors for the effectiveness of anthropomorphism, (2) integrating available research to indicate the most important factors that determine the appropriateness of anthropomorphism, and (3) investigating these factors to gain generalizable knowledge in which cases anthropomorphism is a desirable design feature.

Furthermore, as anthropomorphism does also trigger social responses concerning social categorization, there is a risk that stereotypes of HHI are transferred to robots. The most common way for humans to categorize other humans is gender (Carothers & Reis, 2013), and anthropomorphism seems to be crucial for eliciting this categorization in HRI (Perugia et al., 2022). Robots, especially anthropomorphic ones, therefore have the potential to incorporate gender stereotypes (Chita-Tegmark et al., 2019; Eyssel & Hegel, 2012; Tay et al., 2014). For example, in a recent experiment by Perugia et al. (2022) most anthropomorphic robots were perceived as male, sometimes as female, and rarely as gender-neutral due to their appearance features, thus providing novel evidence for a possible male-bias in robot design. Nowadays, a lot of robots are purposely gendered by appearance features, as it has been postulated that the gender of robots can facilitate the interaction (Eyssel & Hegel, 2012; Powers et al., 2005), especially if the occupational stereotype matches the gender of the robot performing a stereotypically female- or male-associated task. Research has focused mostly on the extent to which the representation of overt gender cues influences HRI in different application domains (Chita-Tegmark et al., 2019; Eyssel & Hegel, 2012; Tay et al., 2014). What has received little attention so far is whether not purposely gendered robots (via name, voice, or surface features) are also accompanied by the association of occupational gender stereotypes. This thesis aims to close this gap by investigating if gender is assigned to gender-neutrally designed robots that match the prevailing occupational stereotype.

Overall, the objective of this thesis can be restated in the light of challenging the effectiveness and desirability of anthropomorphic design in HRI in four studies.

1.1 Anthropomorphism in HRI

The term anthropomorphism is often used interchangeably for two reciprocal phenomena in HRI, the tendency to anthropomorphize and anthropomorphism as a design property of robots (Fischer, 2022). Both aspects influence each other but need to be considered separately in order to understand the mechanisms involved. The tendency to anthropomorphize is a rather stable individual disposition to attribute human characteristics to inanimate objects or animals, in order to rationalize their actions (Waytz et al., 2010). Importantly, it goes beyond animism (i.e., attributing life to nonliving objects), by "attributing capacities that people tend to think of as distinctly human to nonhuman agents, in particular, human-like mental capacities (e.g., intentionality, emotion, cognition)" (Waytz et al., 2010, p. 220). If people anthropomorphize and attribute mind to non-human entities like robots, comparable mechanisms in the brain are activated as when observing the actions of other humans (Wiese et al., 2017). This can widely facilitate HRI from engaging feelings of social connection to enhancing performance in joint tasks (Wiese et al., 2017). Anthropomorphism, as a design feature, aims to trigger this tendency and the resulting consequences for perception and behavior (Fink, 2012; Persson et al., 2000).

Anthropomorphic design can serve as a built-in morphological trigger that activates HHI schemes and expectations. In particular, the physical appearance of a robot is especially relevant for first encounters and the initial interaction (Haring et al., 2016; Hegel, 2012; Phillips et al., 2018), because it has such a high salience (Persson et al., 2000). However, the external appearance of robots is only one option to implement anthropomorphic design features (Fong et al., 2003; Onnasch & Roesler, 2021; Yanco & Drury, 2002). Comparably to the impression forming in HHI illustrated via the quotation at the beginning, movement, communication, and context are also important aspects of the robot's morphology which can be used to induce anthropomorphism (Onnasch & Roesler, 2021). Besides appearance, context can trigger the process of anthropomorphization already before the actual interaction begins (Darling et al., 2015; Kopp et al., 2022; Onnasch & Roesler, 2019). An anthropomorphic context can be induced via framing. This can be done by giving a robot a human name (Keay & Graduand, 2011) or a personified story (Darling et al., 2015). Both, anthropomorphism in the case of context and appearance set expectations for the capabilities of the robot (Haring et al., 2013). As the interaction of humans and robots proceeds, characteristics of the behavior and interaction modalities play a key role in the HRI (Lemaignan et al., 2014). In this process, anthropomorphic communication (Bonarini, 2020) and movement (Kuz et al., 2013; Mayer et al., 2013) can be applied to robots. Both features can be used to convey the intentions of the robot to enhance coordination between both agents (Ferland et al., 2013; Mayer et al., 2013; Riek et al., 2010). Taken together, the morphology of robots in regard to appearance, context, communication, and movement is used to influence the human counterpart on a subjective and behavioral level. Generally, it is assumed that the influence of anthropomorphism is a positive one for the interaction (Breazeal, 2003; Duffy, 2003).

However, as manifold as the ways to implement anthropomorphism by design, as diverse are the outcomes. As stated beforehand, quite generally, it is assumed that a certain degree of anthropomorphism facilitates the HRI due to increased intuitiveness of the interaction (Duffy, 2003; Mori et al., 2012; Złotowski et al., 2014). This only applies to a certain degree of anthropomorphism. If robots too closely resemble humans, the affinity towards robots descent into eeriness (Bartneck et al., 2007; Kim et al., 2020; Mori et al., 2012). Even though there are different specifications for the turning point of the so-called *uncanny valley*, robots have to already highly incorporate anthropomorphic features to fall into it (Mori et al., 2012). One might therefore consider lower levels of anthropomorphism, far away from human replicas, as beneficial. However, this is not always the case for the interaction's human-related outcomes.

1.1.1 The Effects of the Form-Function Relationship

Attributions about the functionality of robots are based on their form and behavior (Lohse et al., 2007), however, these attributions do not necessarily mirror the actual functionality of the robot (Haring et al., 2018). Therefore, negative effects can emerge in the interaction from both breaking and inappropriately shaping expectations. For example, a robot with eyes is expected to visually perceive its environment (Haring et al., 2013) and to show human-like gaze behavior to enable joint attention (Chaminade & Okka, 2013). A robot with randomly moving eyes breaks with these expectations. Moreover, randomly moving eyes seem to inappropriately shape the expectations in the context of industrial HRI, where technological superiority rather than human-like team partnership might be expected (Onnasch & Hildebrandt, 2021; Roesler et al., 2020). Both aspects can lead to a negative effect in comparison to a robot without eyes in regard to perceived reliability (Onnasch & Hildebrandt, 2021; Roesler et al., 2020). Most importantly, a purely decorative anthropomorphic appearance can not contribute to a more intuitive HRI in regard to behavior. In contrast, implementations of anthropomorphic gaze behavior (Moon et al., 2014; Onnasch et al., 2022; Wiese et al., 2018) to robots can facilitate the interaction by triggering the interaction partner's attention to relevant areas.

As the starting point of the research presented here, it was assumed that the effectiveness of anthropomorphism depends on multiple factors like the field of application, the task relevance, and the kind of morphology which is implemented. In addition, it can be supposed that subjective (e.g., perception) and objective (e.g., performance) human-related outcomes do not need to go hand in hand. For example, anthropomorphic appearance can significantly influence human perception (e.g., reliability) but might not significantly change human monitoring behavior in a collaborative task (Onnasch & Hildebrandt, 2021; Roesler et al., 2020). Vice versa, anthropomorphic communication via gaze can significantly influence human behavior in collaborative tasks, however, this does not need to be accompanied by a significant change in perception (e.g., naturalness) (Moon et al., 2014). Anthropomorphism can lead to diverse outcomes on subjective and objective levels. However, current knowledge about these outcomes and the role of contextual factors that make anthropomorphic robot design effective have not yet been systematically identified.

The first paper of this thesis, therefore, aimed to disentangle available research by a comprehensive meta-analytic integration. To provide an overall picture of the current body of research, first, the study postulated a theoretical framework including relevant human-related outcomes and moderators for the effectiveness of anthropomorphism. Second, the study incorporated an extensive systematic literature search, and third, quantitatively combined the respective findings. This meta-analysis aimed to provide insights into the general effectiveness of anthropomorphism, as well as the circumstances under which effectiveness occurs for different subjective and objective outcomes.

1.1.2 The Effects of the Function-Form Relationship

As described beforehand, the appearance and behavior of robots lead to attributions about the functionality of robots. However, this is not a one-way street, as the functionality required for the robot's tasks sets expectations for the appearance and behavior of the robot fulfilling it (Goetz et al., 2003). As robots are making inroads in various tasks of our private and working life, from social support to industrial manufacturing (Onnasch & Roesler, 2019), the question arises of which robot design is preferred for different tasks in different contexts. One of the first to address this question were Goetz et al. (2003). The researchers proposed the *matching hypothesis*, which assumes that the appearance and behavior of robots should match the task and situation. Specifically, they hypothesized "that a more humanlike appearance is a better match for jobs that are more, rather than less, social in nature" (Goetz et al., 2003, p.56). In line with this assumption, the results showed that participants preferred more anthropomorphic robots for more artistic and social jobs (e.g., drawing instructor or tour guide). Conversely, more machine-like robots were favored for jobs categorized as conventional and realistic (e.g., food carrier or security guard). However, the transferability of the results to real robots is not clearly given, as the experiment included customized depictions of robotic heads, which were morphologically limited to facial features and not comparable to actual existing robots.

Building on these results, Li et al. (2010) investigated the matching hypothesis in actual interactions of humans with differently designed LEGO robots, enabling the experience of the complete morphology of each robot. The results, however, did not show a higher preference for robots when the robot's appearance matched the (non-)social nature of the task. That is, an anthropomorphic robot was not significantly preferred as a teacher, and a machine-like robot was not significantly preferred as a security guard. In a post-experiment interview, participants indicated that both the human-like and machine-like robots were most suitable for the security task compared to any other task. This might be related to the Lego Mindstorm NXT robots used for both conditions. Even though the human-like robot differed from the machine-like robot in regard to body shape and eyes, the surface of both robots was highly machine-like. Thus, the surface might have guided the perception of the human-like robot in a more machine-like way.

More recently, Złotowski et al. (2020) investigated the relationship between anthropomorphism and specific tasks with depictions of existing robots. The robots chosen for this experiment, however, differed on various levels beyond anthropomorphism (e.g., threat or likability). If we focus on the two robots, which primarily differed in regard to anthropomorphic appearance, the results showed support for the matching hypothesis. In particular, the more anthropomorphic robot was preferred for more social jobs like child-minder, companion, nurse, or teacher. In contrast, the less anthropomorphic robot was preferred for less social jobs like servant or cleaner.

Overall, the results of the mentioned studies provide some (Goetz et al., 2003; Złotowski et al., 2020), albeit not fully consistent (Li et al., 2010) evidence that the preferred degree of anthropomorphism depends on the task, the robot is expected to fulfill. Furthermore, the results are hardly generalizable to actual HRI due to the robots and tasks used in these experiments. The robots either lacked realism (Goetz et al., 2003; Li et al., 2010), differed on multiple levels beyond anthropomorphism like gender, age, or familiarity (Goetz et al., 2003; Złotowski et al., 2020), or might not have differed in their perceived degree of anthropomorphism (Li et al., 2010; Złotowski et al., 2020). Moreover, the tasks used in all studies (e.g., security guard or tour guide) were rather specific and differed in regard to various aspects like the application domain (e.g., social or service domain), and task specifications (e.g., sociability needed for the task). In addition, none of the studies compromised tasks of the industrial domain, leading to a considerable research gap.

The second and fourth studies of this thesis have been performed to overcome those limitations by systematically taking the application domain and task sociability into consideration. The objective of the second study was to investigate the influence of the application domain (i.e., social, service, and industrial) on the preferred degree of anthropomorphism. Furthermore, the fourth study incorporated the role of task sociability in the social and industrial domain for the preferred degree of anthropomorphism. Taken together, both studies aimed to enable more fine-grained insights into the desirability of anthropomorphism in different domains.

1.2 Gender in HRI

Current research suggests that robots implemented for specific jobs like nursing or security guarding should not only differ in their anthropomorphism (Goetz et al., 2003; Li et al., 2010; Złotowski et al., 2020), but also in their manifested gender (Tay et al., 2014). The idea behind this assumption is that social cues like robot gender are a sufficient anchor to build a common ground about the skills and knowledge of the robot (Powers et al., 2005). Based on occupational gender stereotypes in HHI, humans transfer expectations to gendered robots (Chita-Tegmark et al., 2019; Eyssel & Hegel, 2012; Powers et al., 2005; Tay et al., 2014). The existing occupational gender stereotypes result from different reasons. Different jobs are assumed to require different personal characteristics, which are often associated with gender-trait stereotypes (e.g., women are supposed to be caring) (Koenig, 2018). An existing unbalanced female-to-male ratio in various jobs, and based on this status quo, one binary gender is assumed to be fitting better to the respective job (Adachi, 2013; Lampousaki, 2010). Consequently, it is expected that a fit of occupational stereotype and robot gender can be used consciously to increase acceptance and facilitate the interaction (Powers et al., 2005; Tay et al., 2014).

1.2.1 Intentional Manifestation of Gender

As binary gender manifestation is an important consideration in robotics design (Nomura, 2017), it is a key concern to understand which properties of the robot are used to elicit a gender-specific association. Basically, it can be assumed that all features used to make a robot more anthropomorphic (i.e., appearance, communication, movement, and context) (Onnasch & Roesler, 2019) can be applied in a more or less gendered way (Powers et al., 2005; Robertson, 2010). The most commonly gendered features of robots are names and voices (Bryant et al., 2020; Chita-Tegmark et al., 2019; Jackson et al., 2020; Tay et al., 2014), as well as surface features (Eyssel & Hegel, 2012; You & Lin, 2019), and body proportions (Bernotat et al., 2017, 2021). These features can be used to evoke similar cognitive reactions seen in HHI (Bryant et al., 2020), and consequently aim to make the interaction more intuitive.

The validity of this assumption is at least questionable, as triggering such stereotypical expectations entails both advantages and disadvantages. On part of the advantages, multiple studies illustrated that gendered robots are perceived as more suitable for gender-stereotypical tasks (Bernotat et al., 2017, 2021; Eyssel &

Hegel, 2012). Other study results showed no significant differences between female- and male-gendered robots in the perceived task competence of gender-stereotypical tasks, thought (Bryant et al., 2020; Kuchenbrandt et al., 2014; Lugrin et al., 2020). Yet others showed (Reich-Stiebert & Eyssel, 2017) or postulated (Powers et al., 2005) possible positive effects of mismatching prevailing stereotype and manifested robot gender. So, the question still remains open whether it is generally beneficial to attribute gender and stereotypically accompanied capabilities to robots. The biggest disadvantage of gendered robots is that this novel technology might be reproducing and possibly strengthening old narratives on gender stereotypes (Alesich & Rigby, 2017). Designing robots in a gender-neutral manner might appear to be a suitable counteraction to avoid the transfer of HHI stereotypes to HRI.

1.2.2 Unintentional Transfer of Gender Stereotypes

Currently, it is broadly unknown if using gender-neutral design of robots is a suitable approach to reduce or even eliminate stereotypic associations and reactions on part of the human (Seaborn & Pennefather, 2022a, 2022b). Moreover, it is unclear whether the perception of gender-neutrality is mainly associated with the morphology or also depending on tasks, which might be perceived as more or less gender-neutral (Seaborn & Pennefather, 2022b). Based on HHI it seems reasonable, that not only the morphology of robots (Perugia et al., 2022), but also the existing occupational stereotypes could trigger the perception of robots concerning gender. Gender-based stereotypes about occupations are often based on an assumed link between people's gender and personal characteristics. Most prominently, men are more associated with achievement-oriented characteristics (e.g., being forceful and decisive) and women with social-oriented characteristics (e.g., being helpful and caring) (Heilman, 2001; Koenig, 2018). As mentioned beforehand, one gender seems to be considered as more suitable for an occupation, which is also often reflected in the unbalanced female-to-male ratio in the respective job (Adachi, 2013; Lampousaki, 2010). However, this is not only the case for a specific job, but rather for application domains as a whole. Traditionally, jobs in the industrial domain are rather associated with men, and jobs in the social domain rather with women due to gender-trait stereotypes and the current female-to-male ratio (Heilman, 2001; Koenig, 2018; White & White, 2006). This association might trigger a gendered perception of even gender-neutrally designed robots in the respective domain. Yet, this assumption has not been the focus of HRI research.

The second study of this thesis, therefore, aimed to close this research gap by investigating the influence of the application domain (i.e., social, service, and industrial) on the perceived gender of robots. Most importantly, this was done by using robots without obviously gendered surface features (e.g., hairstyle or lip color) (Eyssel & Hegel, 2012; You & Lin, 2019), and body proportions (Bernotat et al., 2017, 2021).

However, by using robots without obviously gendered features, another challenge emerges. Even though robots might not represent obvious gender cues in appearance, especially the more anthropomorphic robots might be perceived as male-gendered because of various reasons. First, it can be assumed that the association of sophisticated technology like robots and masculinity is present (Faulkner, 2001). Second, robots that are designed without obvious gender cues may still be perceived as male in their appearance. In line with this assumption, a recent study of Perugia et al. (2022) showed that human likeness negatively correlates with gender neutrality. Moreover, it was revealed that the attribution of femininity was related to surface features like eyelashes or hair cuts whereas the attribution of masculinity was attributed to (human) body manipulators like arms or legs. Third, there is another methodological aspect that could also lead to a male perception, apart from the transfer of stereotypes and appearance, namely the language of participants. Many languages assign grammatical gender classes to nouns (Samuel et al., 2019). As the German word robot is grammatically gendered as male, this could have influenced the results of the second study of this thesis and other research investigating stereotypes in HRI via German samples (Ladwig & Ferstl, 2018; Reich-Stiebert & Eyssel, 2017). The assumption that this might lead to a male-gendered perception of gender-neutral robots is supported by a study of Ladwig and Ferstl (2018). They found that most participants in a mostly German sample ascribed a male gender to gender-neutral designed robots. This, however, has not yet been brought into the context of language.

The goal of the third study of this thesis was to find out which robots are perceived as gender-neutral in their appearance. This was done as the robots without obvious gender cues used in the second study might have been perceived as rather non-feminine robots than really gender-neutral ones. In addition, the study aimed for a comparison between grammatically gendered and non-gendered languages to answer whether the gendered perception of robots is language specific. Moreover, this study investigated whether the direct assignment of grammatical gender to inanimate objects like robots leads people to think of them in a gendered manner (Boroditsky et al., 2003). This study thus also offered the opportunity to select robots that are clearly perceived as gender-neutral in their appearance to further investigate a possible transfer of gender stereotypes to gender-neutral robots.

The objective of the fourth study was therefore to investigate the possible transfer of occupational gender stereotypes to gender-neutral robots. The goal was to find out if gender-neutral robots offer the opportunity to eliminate stereotypical associations or are still prone to stereotypes of HHI.

1.3 The Approach of the Present Research

As discussed in section 1.1.1, much research had been conducted to investigate the effects of anthropomorphism in HRI (Fink, 2012; Fischer, 2022; Złotowski et al., 2014). Most researchers postulated advantages of applying a certain degree of anthropomorphic features to robots, especially in the social domain (Breazeal, 2003;

Duffy, 2003; Fong et al., 2003). Others illustrated downsides in predominantly industrial settings (Onnasch & Hildebrandt, 2021; Roesler et al., 2020). Despite the relevant implications for the specific interaction scenarios, general knowledge under which circumstances this design approach is beneficial, ineffective, or even detrimental is still missing. The first study of this thesis aimed to close this gap by applying meta-analytic methods to the current body of research on anthropomorphism in HRI. This approach was chosen because it seemed to be most appropriate to summarize the already available knowledge and empirical evidence in the field. In addition, this approach was used to evaluate which possible blind spots in the field are present and where more experimental work is needed.

Based on the results of the meta-analysis that anthropomorphism seems to be effective only in certain application domains, the online studies of this thesis were conducted. Two of these studies aimed to answer the question in which application domains anthropomorphism is desirable due to a possibly fitting function-form relationship. Moreover, anthropomorphism might be generally not desirable due to a possible replication of existing occupational gender stereotypes (Alesich & Rigby, 2017; Heilman, 2001; Koenig, 2018; White & White, 2006) to even gender-neutral robots. However, to the best of my knowledge, the relationship between anthropomorphism and gender-neutral robots in regard to occupational stereotypes was not yet investigated. Therefore, three online studies were conducted to investigate the preference for anthropomorphic appearance (section 1.1.2) and gender ascription in different domains (section 1.2.2). This approach was chosen as, in light of the rapid spread of COVID-19, research had to broadly move online (Feil-Seifer et al., 2020). Nonetheless, as the general attitudes of the public and not behavior towards the robot were of main interest, an online approach seemed to be reasonable. In particular, the second study of this thesis investigated whether different application domains are associated with a different preference for anthropomorphism. In addition, gender attributions of robots were investigated via a naming technique to allow for attributions apart from binary ones. That is, the naming technique gave participants the possibility to not only assign traditional male or female names but also nicknames and functional names. The third study addressed possible methodological drawbacks of the second study by considering the appearance of robots without obvious gender cues and the role of grammatical gender. The knowledge gained in this study allowed the fourth study to incorporate clearly gender-neutral robots to examine the role of the application domain and task sociability for both preferred anthropomorphism and associated gender.

In conclusion, this thesis aimed to shed light on the effectiveness and desirability of anthropomorphism in HRI. Most of all, the incorporated studies aimed to broaden the view beyond isolated and rather specific interaction scenarios, and included social issues possibly emerging from anthropomorphic robot design.

2 Publications

Publication 1: Roesler, E., Manzey, D., & Onnasch, L. (2021). A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction. *Science Robotics*, 6(58), https://doi.org/10.1126/scirobotics.abj5425.

Publication 2: Roesler, E., Naendrup-Poell, L., Manzey, D., & Onnasch, L. (2022). Why context matters: the influence of application domain on preferred degree of anthropomorphism and gender attribution in human–robot interaction. *International Journal of Social Robotics*, 14, 1155–1166, https://doi.org/10.1007/s12369-021-00860-z.

Publication 3: Roesler, E., Heuring, M., & Onnasch, L. (submitted). (Hu)man-like robots — the impact of anthropomorphism and language on perceived robot gender. Submitted to *International Journal of Social Robotics – Special Issue on on Gendering Robots (GenR): Ongoing (Re)configurations of Gender in Robotics International Journal of Social Robotics.*

Publication 4: Roesler, E., Heuring, M., & Onnasch, L. (submitted). Robots in different jobs — Associations of anthropomorphism and gender stereotypes. Submitted to *Technology in Society.*

2.1 Publication 1: A Meta-Analysis on the Effectiveness of Anthropomorphism in Human-Robot Interaction

Eileen Roesler, Dietrich Manzey & Linda Onnasch

Document type: Journal article | Accepted version (i.e., final author-created version that incorporates referee comments and is the version accepted for publication; also known as: Author's Accepted Manuscript (AAM), Final Draft, Postprint). This is the author's version of the work. It is posted here by permission of the AAAS for personal use, not for redistribution. The definitive version was published in Science Robotics on Vol 6, Issue 58 from September 8th 2021, DOI: 10.1126/scirobotics.abj5425.

Availability: Open Access via institutional repository of Technische Universität Berlin via https://doi.org/10.14279/depositonce-12447

Eileen Roesler, Dietrich Manzey, Linda Onnasch

A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction

Abstract

The application of anthropomorphic design features is widely assumed to facilitate human-robot interaction (HRI). However, a considerable number of study results point in the opposite direction. There is currently no comprehensive common ground on the circumstances under which anthropomorphism promotes interaction with robots. This meta-analysis aims to close this gap. A total of 4,856 abstracts were scanned. After an extensive evaluation, 78 studies involving around 6,000 participants and 187 effect sizes were included in this meta-analysis. The majority of the studies addressed effects on perceptual aspects of robots. In addition, effects on attitudinal, affective, and behavioral aspects were also investigated. Overall, a medium positive effect size was found, indicating a beneficial effect of anthropomorphic design features on human-related outcomes. However, closer scrutiny of the lowest variable level revealed no positive effect for perceived safety, empathy, and task performance. Moreover, the analysis suggests that positive effects of anthropomorphism depend heavily on various moderators. For example, anthropomorphism was in contrast to other fields of application, constantly facilitating social HRI. In conclusion, the results of this analysis provide insights into how design features can be used to improve the quality of HRI. Moreover, they reveal areas in which more research is needed before any clear conclusions about the effects of anthropomorphic robot design can be drawn.

Introduction

Robots are making inroads into our working life and everyday world (*1*, *2*). Whereas early robot generations were mainly limited to industrial robots that worked in safety cages, kept apart from human workers, current robotic agents are increasingly interactive. In this process, interaction is changing from a segregated coexistence to direct collaboration with humans in the same space and time. The ability to collaborate, in turn, enables the implementation of robots in more diverse domains (*3*). In addition to being deployed in industrial settings, robots are also becoming more common in service and social fields of application such as school teaching and elderly care. This general shift of robots entering the world of humans is increasingly accompanied by the application of human-like features in robot design (*4*–*7*). The postulated effectiveness of this anthropomorphic design approach is mainly based on two assumptions. First, robots are used in an environment that is designed and optimized for humans. For this reason, the application of human-like design is assumed to support a naturalistic and functional embodiment (*4*). Structural and functional similarities e.g., limbs and joints provide the capabilities, which can support a successful movement through an environment and an interaction with artefacts built for humans (*8*, *9*). Second, from a human-centered point of view, anthropomorphism promotes more intuitive interaction for people because it enables the transfer of scripts that are well known from human-human interaction (*10*, *11*).

Anthropomorphism in HRI is thereby a reciprocal phenomenon. On the one hand, it describes the general tendency of people to attribute human characteristics including human-like mental capacities to non-living objects (*12*, *13*). On the other hand, anthropomorphism describes a human-like design of robots that in turn facilitates the attribution of human-like characteristics to the robot (*3*). This design element is used to evoke expectations, which, if met, represent a knowledge base for interaction and a better anticipation of robots' actions, even for first encounters with this often completely new technology (*5*, *11*, *14*). Figure 1 shows a number of examples of anthropomorphic robot designs in different domains of human-robot interaction (HRI). The examples also illustrate that most straightforward approaches of anthropomorphic robot design address the overall appearance of robots (e.g., face-like characteristics or body shapes). However, other approaches include more subtle aspects such as anthropomorphic trajectories, language-based communication, or simply different types of framing (e.g., giving robots human names or human-like descriptions).

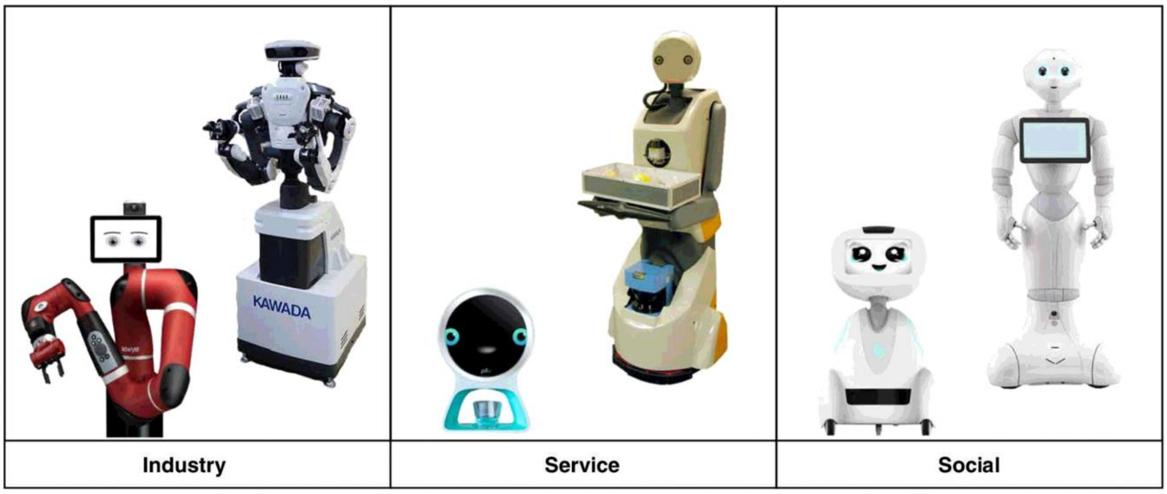

Fig. 1. Examples of anthropomorphic implementations. Anthropomorphic design by means of depicting human-like facial features or body features for the industrial (left: Sawyer; right: Nextage), service (left: Pillo Health ; right: SnackBot), and social domain (left: BUDDY; right: Pepper) received from the Anthropomorphic Robot (ABOT) Database (*15*)

But is this design approach generally beneficial for HRI? While current research in social application domains broadly supports this assumption (*4*, *5*, *12*), a different picture emerges in other domains. For example, studies focusing on industrial HRI suggest that anthropomorphic design features may not necessarily be beneficial, and can undermine the perceived reliability of robots (*16*) and raise concerns with regard to their safety (*17*). These results are unexpected, because the transfer of human-human interaction scrips should make interaction more familiar and trustworthy, independent of the application domain in question.

Interestingly, negative effects are not only observed in the industrial domain, but also in other domains where humans have to perform a certain task in collaboration with a robot. In this case, an anthropomorphic robot representation may again lead to counterproductive and unintended effects, including a decrease in prosocial behavior (*18*), or overshooting effects such as an inappropriately strong emotional attachment to the robot (*19*) .

Overall, these examples suggest that anthropomorphic design can lead to diverse and unintended outcomes. However, our current knowledge about the context factors that make anthropomorphic robot design beneficial have not yet been systematically identified, and a comprehensive integration of the available research is lacking.

With this meta-analysis, we aim to close this research gap by (1) estimating the overall effect of anthropomorphism on human-related outcomes, (2) separately estimating effects of anthropomorphism on different facets of human-related outcomes, and (3) taking into account possible moderators. The basic framework for this analysis, depicted in Figure 2, includes and arranges the key variables considered in our meta-analysis.

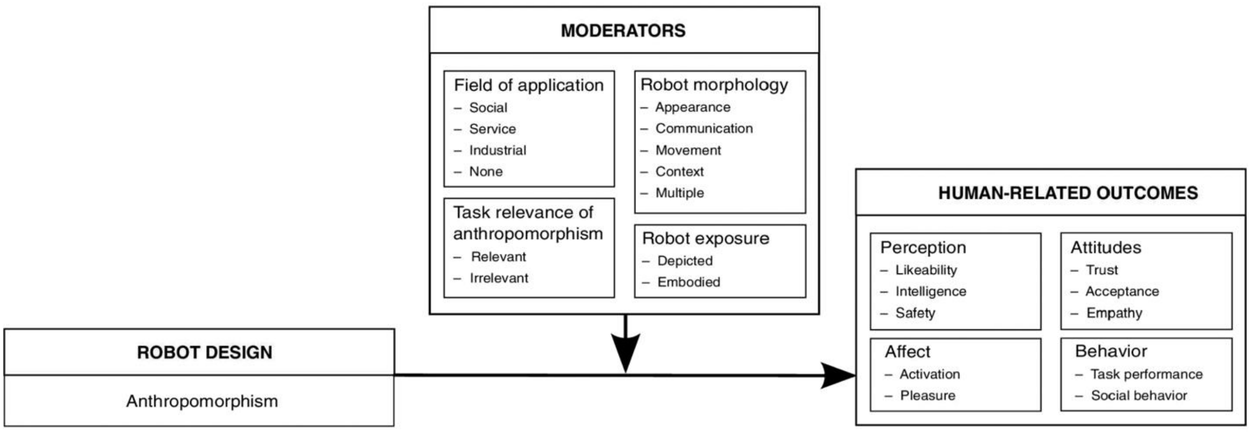

Fig.2. Basic framework of the meta-analysis.

The anthropomorphism of the robot represents the relevant input variable. For this reason, only studies that investigated the effects of at least two different degrees of anthropomorphic robot design were considered in this meta-analysis to estimate the effectiveness of increasing the anthropomorphism of robots. The primary aim of the analysis was to examine the generally assumed positive effects of anthropomorphic robot design. We therefore excluded studies that explicitly address what is commonly referred to as the uncanny valley effect in HRI, which focuses on negative consequences of highly anthropomorphic designs in terms of disturbance and eeriness (*20*).

The relevant dependent variables are summarized as human-related outcomes in terms of subjective and objective interaction experiences (*21–24*). We identified four main categories of outcomes based on an extensive analysis of the current body of research. The first category is people's perception of robots. Most of the relevant research in this area was based on the Godspeed questionnaire series (*25*). Besides evaluating anthropomorphism and animacy itself, this questionnaire series assesses how likeable, intelligent and safe a robot is perceived to be by the human counterpart. The second category covers different attitudes towards robots. Previous research has shown that attitudes such as trust, acceptance, and empathy are important determinants of people's actual behavior in HRI, and specifically their willingness to work together with their robotic counterpart (*26*, *27*). Whereas trust (*26*, *27*) and acceptance (*28*, *29*) are assumed to be mainly associated with effective and efficient interaction, empathy seems to be especially relevant in social HRI settings (*22*, *30*). The remaining two outcome categories include affective reactions (*31–33*), i.e., activation and pleasure in terms of pleasure-arousal theory (*34*, *35*), and behavioral responses, including task performance (*36*, *37*) and social behavior shown in interaction with a robot (*18*, *22*).

To investigate the circumstances under which anthropomorphism facilitates HRI, our analysis further considers several moderating variables. Based on reviews (*14*, *38*) and a recent taxonomy of HRI (*39*), we identified four central moderators that might explain possible heterogeneity in individual study results. The first moderator relates to the interaction environment, and sets the conditions and constraints for the configuration of interaction, i.e., the field of application (*39*). The fields of application considered are categorized as the social, service, and industrial domain. The social domain is defined as any domain where robots are used in therapeutic, educational, or entertainment settings (*39*). The service and industrial domain are defined based on the International Organization for Standardization and (ISO 8373:2012) (*40*). In these fields of application, robots perform useful functional tasks for humans such as transport, physical load reduction, and precision. In addition, this moderator variable includes a fourth category ("none"), given that some HRI studies focus on the pure perception of robots without any contextual information.

The next two moderator variables include different aspects of the robot itself. One is the instrumentality of the anthropomorphic design feature. Studies suggest that it might make a difference whether or not anthropomorphic features are related to the task in a meaningful manner (e.g., randomly moving eyes vs. predictive eyes (*41*)). Whereas task-relevant implementation may lead to increased task performance, this is probably less the case with task-irrelevant implementation of anthropomorphic features. In addition, the impact of this moderator might also be different for various outcome categories of HRI. In contrast to task-irrelevant implementations, task-relevant anthropomorphic design might directly improve actual performance, but it seems less obvious whether it also differently affects people's perception of or attitude toward robots. The third moderator addresses how anthropomorphic features are implemented in the robot's morphology (*39*), i.e., the appearance, communication, movement and/or the context in which the robot is framed and introduced to users. We assume that different implementations of anthropomorphism can be variously effective with regard to different outcome categories. For example, whereas an anthropomorphic appearance might not affect task performance (*42*), anthropomorphic movements might do so by improving predictability of the robot's actions, thereby enhancing coordination in task fulfilment (*43*). In addition to the four implementation categories, a fifth is added to cover cases where multiple anthropomorphic features are combined.

Finally, the last moderator in the framework comprises a more research-relevant aspect, involving the question of how to expose humans to robots in HRI studies, i.e., whether humans interact directly with embodied robots (i.e., real machines) or must merely imagine interaction based on depictions of robots (i.e., virtual two-dimensional agents). Both approaches are used in HRI research, but there is no comprehensive ground yet regarding how this might affect the results (*44–46*). To shed light on this issue, robot exposure is included in this analysis by categorizing the robots used as either depicted or embodied (*39*).

In summary, although consequences of anthropomorphic features in HRI have been investigated widely, we still lack knowledge about the generalizability of specific results produced by individual studies. Based on the proposed framework, this study aims to systematically review and quantify the effects of anthropomorphism on identified human-related outcomes. Moreover, the analysis takes into account the role of moderators to enable a differentiated understanding with regard to the circumstances under which anthropomorphism can facilitate or hinder HRI. To achieve this goal, we applied quantitative meta-analytic methods to the existing literature on anthropomorphism in HRI.

Results

Figure 3 illustrates the overall effect of anthropomorphism, as well as the effects for the different outcome categories and specific variables.

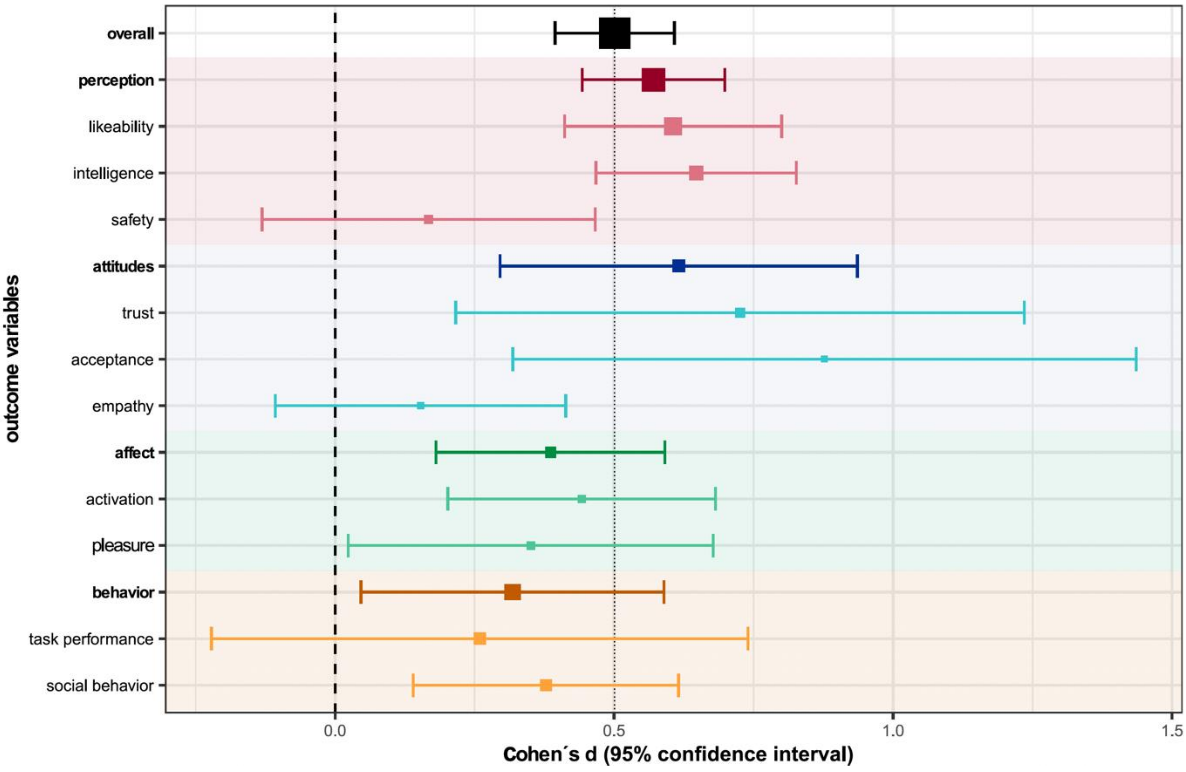

Fig. 3. Forest plot of the overall effect size and all sublevels. Depiction of standardized mean differences (Cohen's d) shown by the positions of the squares, the 95% CIs by the whiskers, and the numbers of included studies by the size of the squares.

Overall effect

The analysis revealed a positive overall effect of anthropomorphism on human-related outcomes with a medium average effect size (*d*=0.501, 95% CI [0.394-0.608]). However, the analysis also revealed a high level of heterogeneity ($Q_{(186)}$= 1684.25, p<.001, I^2=88.1%), suggesting diverse effects on different outcome variables and/or an impact of moderator variables.

Human-related outcomes

Perception. The analysis showed that people's perception of robots is the most frequently investigated construct (*k*=99) to evaluate the consequences of anthropomorphic design in HRI. Overall, the reported effects of anthropomorphism on perception result in a medium average effect size (*d*=0.570, 95% confidence interval (CI) [0.443-0.698]), again with a high level of heterogeneity ($Q_{(98)}$= 753.57, p<.001, I^2=84.93%). The separate analyses for the different subdimensions suggest that the overall positive effect of anthropomorphism on people's perception of the robot is mainly driven by the subcategories of likeability (*d*=0.606, 95% CI [0.411-0.800]) and intelligence (*d*=0.647, 95% CI [0.467-0.827]). In contrast, the data revealed no consistent effect for studies addressing the perceived safety of robots (*d*=0.168, 95% CI [-0.131-0.466]).

Attitudes. A similar pattern of effect sizes emerged regarding attitudes towards robots, although this aspect was based on a considerable smaller set of studies (*k*=25). The analysis again revealed a positive overall effect (*d*=0.616, 95% CI [0.296-0.936]) with a pronounced heterogeneity ($Q_{(24)}$= 199.80, p<.001, I^2=90.51%). The subset analyses showed that the overall effect was mainly due to two subcomponents, i.e., a positive effect of anthropomorphism on trust with a medium effect size (*d*=0.726, 95% CI [0.216-1.235]), and a positive effect on acceptance with a large effect size (*d*=0.877, 95% CI [0.318-1.436]). In contrast, no consistent positive effect was found for empathy towards robots (*d*=0.153, 95% CI [-0.107-0.413]).

Affect. The effects of anthropomorphism on affect are least investigated, having been addressed in only *k*=18 studies. The mean effect size of these studies is again positive (*d*=0.386, 95% CI [0.181-0.591]). Compared to the effects on perception and attitudes, it is somewhat smaller, but also more consistent with less remaining heterogeneity ($Q_{(17)}$= 37.58, p<.01, I^2=55.67%). In this case, the overall effect is also representative for both subcomponents, characterized as activation (*d*=0.441, 95% CI [0.202-0.682]) and pleasure (*d*=0.351, 95% CI [0.023-0.678]).

Behavior. The effects of anthropomorphism on human behavior in HRI were addressed in k=45 studies. Overall, anthropomorphism has a small positive effect on this outcome category (d=0.318, 95% CI [0.046-0.590]). This positive effect can be mainly traced back to beneficial effects on social behavior (d=0.378, 95% CI [0.140-0.616]). In contrast, no consistent improvements emerged for task performance (d=0.259, 95% CI [-0.222-0.740]). In line with the results for perception and attitudes, the analysis of this outcome category also revealed a large degree of systematic heterogeneity between studies ($Q_{(44)}$= 616.10, p<.001, I^2=91.68%), again suggesting the effects of moderator variables.

Moderators

The results presented show that the meta-analytic models used to analyze the effects of the different studies almost always indicated a relatively high level of heterogeneity in the data. This suggests that moderators most likely contributed to the differences between studies. Figure 4 shows the results of the moderator analyses addressing the set of a priori identified outcome categories. Each single graph in the figure illustrates the differences between mean effect sizes dependent on the categories of a given moderator (columns) and the different outcome categories (rows).

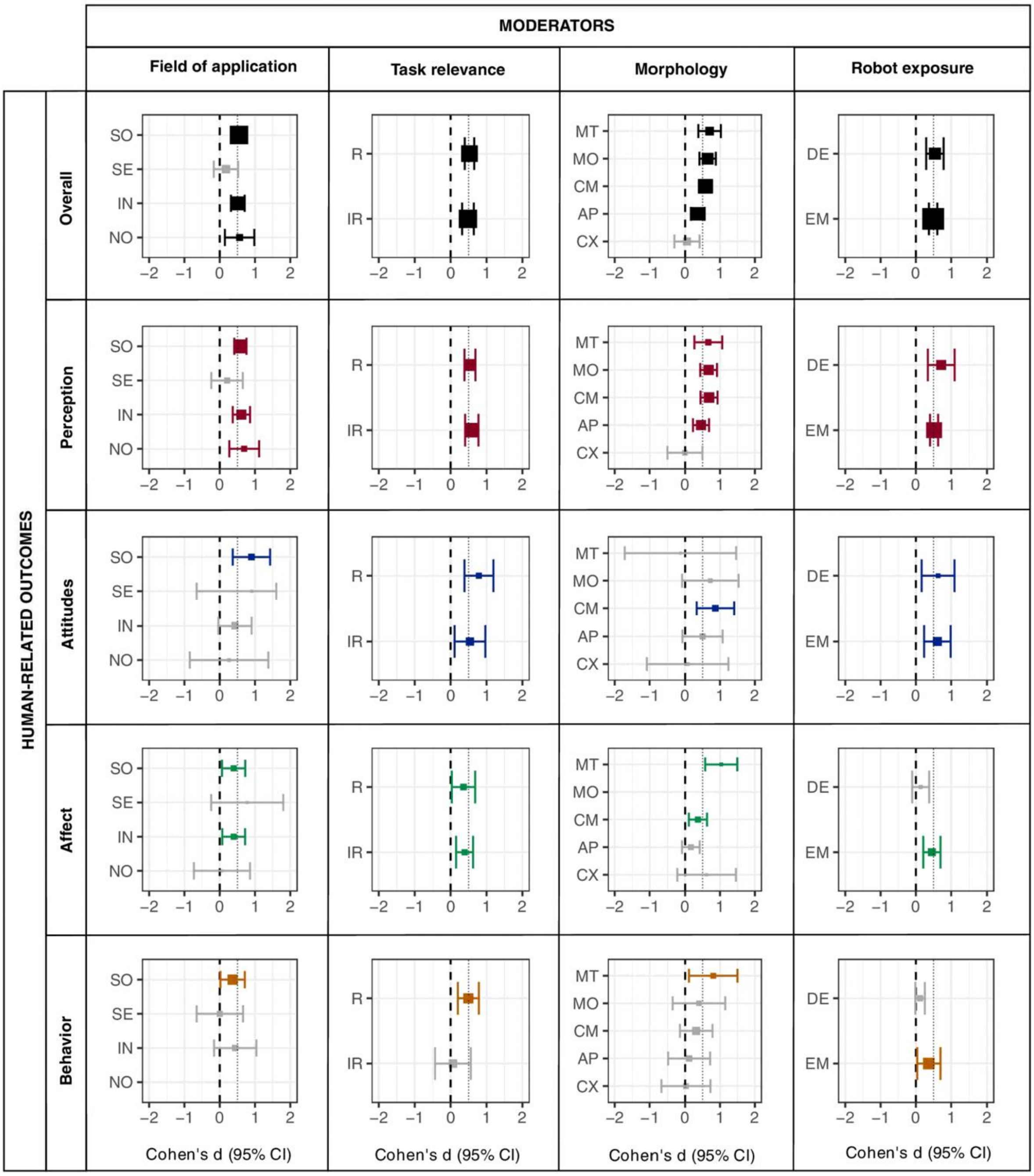

Fig. 4. Forest plots illustrating the effects of moderators. The plots show standardized mean differences (Cohen's d), the 95% CIs, and the number of effect sizes included, given separately for the overall effect and all subcategories, dependent on the characteristics of the different moderators (columns). The moderator variables are (i) field of application (SO, social; SE, service; IN, industrial; NO, none), (ii) task relevance (R, relevant; IR, irrelevant), (iii) morphology (MT, multiple; MO, movement; CM, communication; AP, appearance; CX, context), and (iv) robot exposure (DE, depicted; EM, embodied).

Field of application. On an overall level, the field of application explained only 0.9% of heterogeneity (Q_M= 3.83, p=.28). Closer scrutiny reveals that a consistent positive effect size across all different outcome categories was only found for the social domain, whereas no comparable consistent effects of anthropomorphism emerged for studies of HRI in the service domain. A somewhat mixed pattern of results emerged for the industrial domain. In this case, anthropomorphism yielded small to medium effects for perceptual and affectional outcomes. Finally, studies with no clearly defined field of application found consistent beneficial effects of anthropomorphism for people's perception of the robot only, while no comparable consistent results were found for the other sets of outcome categories.

Task relevance. For the overall effect, task relevance did not account for any heterogeneity (Q_M= 0.28, p=.597). Independent of whether or not anthropomorphic design features were implemented in a task-relevant manner, they led to positive effect sizes for all outcomes apart from behavioral ones. For this latter category, the task relevance of anthropomorphic features seems to be a necessary condition for achieving positive effects.

Morphology. The overall positive effect of anthropomorphism was moderated by how anthropomorphism was implemented, i.e., the dimension used to increase the anthropomorphism of a robot (Q_M= 11.44, p<.05). Specifically, multiple implementations (d=0.703, 95% CI [0.38-1.025], p<.01), implementations via movement characteristics (d= 0.645, 95% CI [0.41-0.879], p<.01), and implementation of human-like communication (d=0.583, 95% CI [0.396-0.769], p<.05) significantly increased the positive effect compared to using context framings only (d=0.054, 95% CI [-0.306-0.414]). Regarding appearance, at least a non-significant trend for increased effectiveness compared to the context was found. On the sublevel of outcome categories, communication and multiple implementations of anthropomorphic features most consistently led to positive effects for three of the four outcome categories. Anthropomorphic appearance and movement only resulted in a positive effect size for perception, and the anthropomorphic context did not lead to any positive effect on any of the outcome categories.

Robot exposure. The physical presence of the robot did not account for any heterogeneity (Q_M= 0.099, *p*=.753) of the overall effect. Medium effect sizes with similar values were present for both studies using depicted robots and studies using embodied robots. On the sublevel of the outcome categories, a double-edged picture emerged. Whereas studies using embodied robots report consistent beneficial effects across all outcomes, studies using only depicted robots for their research merely found a positive effect with respect to perception and attitudes.

Publication bias

The visual inspection of the data via a funnel plot showed a left-sided asymmetry, which indicates that more effect sizes were included in our analysis that underestimate the true effect compared to effect sizes that overestimate it. This asymmetry was supported by a significant Egger's regression test for funnel plot asymmetry (*z*=4.47, *p*<.001). More precisely, the trim-and-fill method revealed that the estimated number of missing studies was 26 on the right side and none on the left. In comparison to the uncorrected overall effect of anthropomorphism (*d*=0.501, 95% CI [0.394-0.608]), the trimmed-and-filled dataset resulted in a slightly higher overall effect size (*d*= 0.655, 95% CI [0.542-0.7679]).

Discussion

The objective of this meta-analysis was to investigate the effects of anthropomorphic design features on human-related outcomes, and to take into account relevant moderators. The results reveal that adding anthropomorphic features to HRI leads to a considerable overall positive effect, which is in line with previous research (*21–24*). Moreover, the results show that this holds true for all different outcome categories considered in this analysis, with moderate effects of anthropomorphism for perception and attitudes, and relatively smaller effects for affect and behavior. The analysis further revealed that most studies thus far have focused on the impact of anthropomorphism on perceptual aspects such as the perceived intelligence or likeability of robots (*25*). Thus, the perceptual category represents the most important source for the overall positive effect. This overrepresentation of perception compared to other categories in HRI research does not seem to be justified by its greater relevance. Instead, it seems to be primarily related to the ease of accessibility of this sort of outcome variable. For instance, one of the most commonly used tools in HRI research is the Godspeed questionnaire series (*25*) (and the according revised version (*47*)). This is a very cost- and time-effective measure that addresses aspects of how people perceive robots (*25*, *48*). In contrast, effects of anthropomorphism on affect or even behavioral outcomes require more complex assessment approaches. However, attitudes are also less commonly investigated. This is surprising for two reasons. First, the ease of accessibility of attitudes as a subjective measure is comparable to that of perceptual evaluations (*28*). Second, the positive effects of anthropomorphism on trust and acceptance are some of the most commonly mentioned ones in the literature (*12*, *27*). Obviously, there is a gap between the theoretically postulated importance of attitudes for a successful HRI and gaps in the research on this specific topic that need to be closed by future studies. In addition, our results call for more research on behavioral outcomes. Regardless of the domain in which humans and robots collaborate, the primary goal of anthropomorphic design features will always be to improve behavior (e.g., physical stimulation in therapeutic settings or smooth joint manipulation of work pieces in industry). Of course, it is important to investigate subjective perceptions of robots and attitudes towards them in HRI research (*26*, *27*), given that both presumably determine people's behavior and willingness to work together with a robot. However, actual behavior will always be the key concern, and should not be neglected in research.

More detailed analysis on the specific variable level (per outcome category) further suggests that anthropomorphic design features have no impact on the empathy towards robots, the perceived safety of robots, or performance in joint tasks with robots. The non-existent positive effect of empathy might again be related to the underrepresentation of research on this rather specific aspect (k=7). In contrast, the missing effects on perceived safety and task performance can certainly be considered a reliable finding because the analysis was based on a relatively higher number of studies, specifically in non-social HRI settings. The lack of evidence for improved task performance challenges the assumption that equipping robots with anthropomorphic features might activate human-human interaction schemes in HRI, which, then, intuitively supports task-related behavior (*11*). Combined with the overall null effect on perceived safety, it suggests that anthropomorphic design features might primarily be used to improve social aspects in HRI (*5*, *12*), but not task-related aspects.

The additional consideration of possible moderators generated further insights into the specific circumstances that might determine the effectiveness of anthropomorphism. The first moderator was the field of application. In line with an already sound body of research (*4*, *5*, *12*), the results show that the social domain consistently benefits from the application of anthropomorphism. This positive effect is not directly transferable to other domains, though. Specifically, the service domain does not seem to benefit at all from anthropomorphic robot design. A possible explanation could be that anthropomorphic features lead to an emotional attachment (*19*), which might undermine a person's willingness to use the robot as tool. Whereas anecdotal evidence (*49*) for this assumption exists (e.g., delivery robots are used less if they are anthropomorphized more), further research is needed to consolidate this hypothesis.

The second moderator addressed whether or not it makes a difference if anthropomorphic design features directly relate to the task at hand. Our data confirm the expectation (*41*) that the task relevance of implemented anthropomorphic design features is only a crucial factor for facilitating HRI with respect to behavioral outcomes. This finding seems to be particularly important for actual work-related collaborative interactions. It suggests that it is worthwhile to implement anthropomorphic features in a task-relevant manner (e.g., social cues, predictive movements) whenever humans and robots collaborate on certain tasks.

The third moderator considered in our analysis included effects of how specifically anthropomorphic features were implemented, i.e., based on appearance, the communication channel, movements, or just the type of framing. The data demonstrate that different implementations of anthropomorphism can lead to a variety of effects. Not surprisingly, approaches based on multiple as well as communicational anthropomorphic features turned out to be most effective with regard to the different outcome categories. In contrast, the mere use of different sorts of framing to induce an anthropomorphic context, e.g., giving a robot a name and a personalized story (*18*, *22*), does not seem to be effective, having no reliable overall effect on any of the outcome categories. There may be two reasons for this missing positive effect of context anthropomorphism: the limited salience in comparison to more visually detectable anthropomorphic features, and the possible masking of the robot's functional value by covering its task-related features as a tool (*18*). Other morphological features were effective for some, but not all human-related outcomes. On the one hand, the positive effect of appearance on perception is not surprising, because an anthropomorphic appearance is described as the most salient characteristic (*12*, *21*). On the other hand, it might be possible that anthropomorphic appearance had no effect on attitudes, affect and behavior because of the non-functional character of appearance (*39*). In addition, appearance can establish certain expectations regarding the robot's functionalities that might get violated in following interactions.

Finally, the last moderator variable addressed methodological issues of HRI studies and investigated whether the efficiency of anthropomorphism depends on how the robots are presented to participants, i.e., in a physically embodied manner that allows for lively observation or even direct interaction, or merely by two-dimensional representations. Here, our results reveal a gap between subjective and objective outcomes. Regardless of how participants are exposed to a robot, positive effects of anthropomorphism emerged for perception and attitudes, both of which are usually assessed via subjective questionnaires. However, positive effects on affect and behavior, which concern actual physiological and behavioral reactions (*21–24*), are usually only found in studies that involve presenting "real" robots to the participants. Earlier research indicated both similarities (*45*) and differences (*44*, *46*) between physically embodied robots and virtual two-dimensional representations. The gap between subjective and objective reactions indicates a possible systematic explanation for these mixed results and could be instructive for future research. If perceptions or attitudes towards (anthropomorphic) robots are of the main interest, it seems sufficient and ecologically valid to conduct studies using virtual agents or images of robots. However, if affective or behavioral outcomes are central to an investigation, researchers should seek to use studies involving physically present robots that enable real interaction so as to gain valid insights.

Overall, the analysis suggests that it is counterproductive to draw general conclusions on the impact of anthropomorphism on HRI when these are based solely on perceptual evaluations. Apart from a handful of exceptions (i.e., in the service domain, implemented via the context), anthropomorphism is always beneficial to people's perception of robots. However, this effect does not seem to be transferable to other more reciprocal interactional outcomes such as the behavioral outcomes considered in our meta-analysis. Moreover, the analysis illustrates another even more important issue regarding the transferability of effects of anthropomorphism. Based on the shift of the robot's role from a tool to a team partner (*39*), it has often been assumed that the results gained in social HRI can be transferred to other fields of application. However, the results suggest that the stable positive effect of anthropomorphism in social HRI may not be directly transferable to other domains. For example, essentially no positive effects of anthropomorphism were found in the service domain, and only partial effects were determined in the industrial domain. This shows the inadequacy of transferring insights from social HRI to more task-related settings. Furthermore, the overall effectiveness of anthropomorphism on social behavior, but not on task performance, challenges the usefulness of anthropomorphic features in those domains. In sum, even though the analysis showed no evidence for a negative impact of anthropomorphic design, anthropomorphism also does not generally improve the quality of HRI. Whereas social HRI consistently benefits from anthropomorphic robot design, a mixed picture emerges for other application domains. In addition, the way anthropomorphism is implemented seems to determine its success. Most of all, our results suggest that interaction quality between humans and robots can particularly be promoted by implementing anthropomorphic communication features, by multiple implementations of anthropomorphism and by implementing task-relevant anthropomorphism.

Meta-analyses must always be interpreted with caution, because they equally include measures of various study designs involving different numbers of participants. However, given the systematic procedure and the comparably high number of effect sizes included, we assume that the global conclusions presented above are indeed reliable findings. Moreover, the analysis of possible publication bias suggests that if a bias is present at all, it has biased our analysis conservatively with regard to the impact of anthropomorphic robot design. Nonetheless, one major limitation of the study concerns the non-consideration of different degrees of anthropomorphism. Most of the empirical effects included in the analysis contrasted only two different degrees of anthropomorphism, which could hardly be located on an overall dimension. The main reason for this limitation is that the exact degree of anthropomorphism of robots cannot be measured objectively. Thus, even though it was possible to detect some major moderating factors of effects of anthropomorphism, we are unable to make any conclusions about the degree of anthropomorphism required to induce certain effects (*25*, *48*). This will be a matter of future research, and we hope that our meta-analysis will be a good starting point for such research. The fact that we have the entire data and material of this meta-analysis available online will enable other researchers to add more data and to expand this data base over time. By taking this approach, our meta-analysis serves not only as a state-of-the-art research synopsis, but moreover aims to iteratively create a sound basis for investigating the consequences of anthropomorphism in future science and practice.

Materials and methods

Before starting the systematic literature search, the meta-analysis was preregistered and described in detail in the standardized procedure of preferred reporting items for systematic review and meta-analysis protocols (*50*) via the open science framework (*51*). The entire methodical procedure and all data generated during the process, from the literature search to the actual analysis of the data, are available online to enable other researchers to replicate and further extend the analysis in the future (*51*).

Based on the objective of the study, the terms used for the literature search included combinations of <human-robot interaction or social robot> and <anthropomorphism or anthropomorphic or humanlike> and <experiment or subject or participant or user study)>. The literature search was conducted between April and June 2020. The comprehensive procedure, encompassing also the list of inclusion criteria, is illustrated in Figure 5.

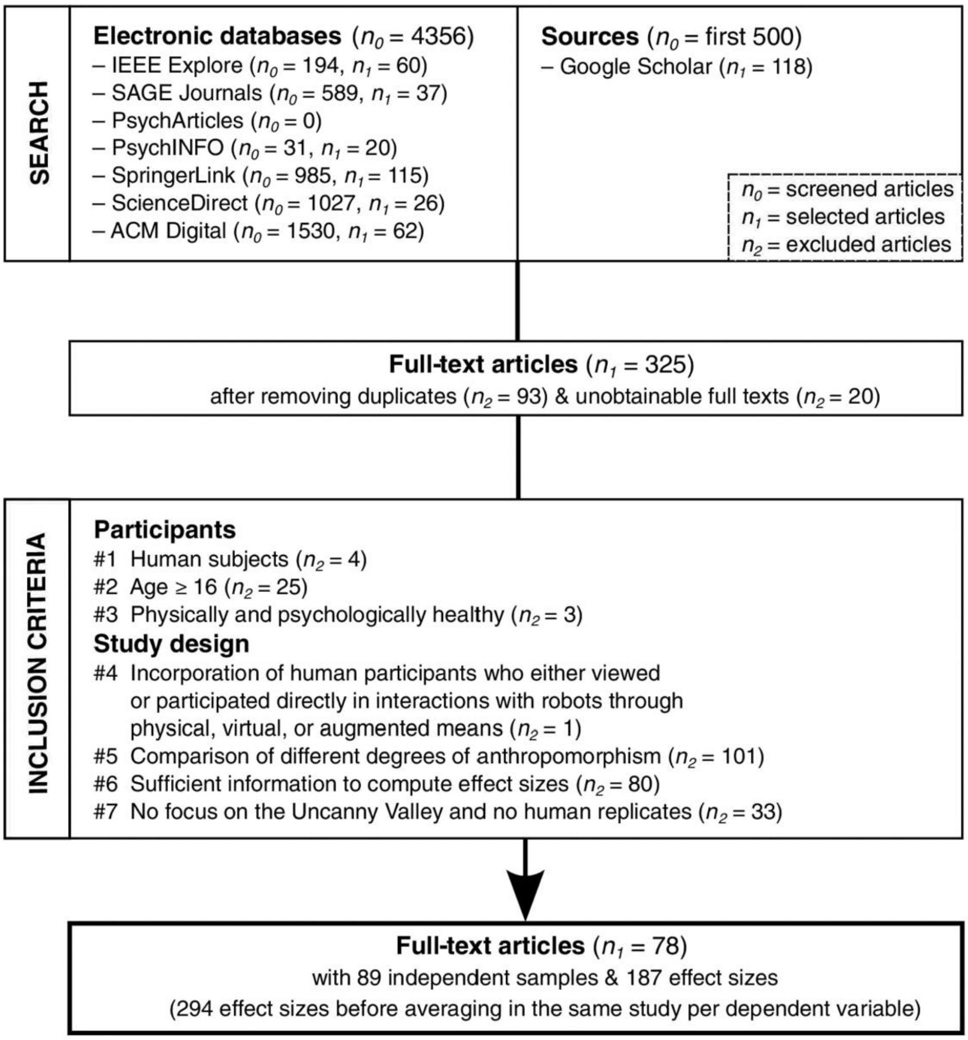

Fig. 5. Search flow diagram. Depiction of the entire process of data collection, including the sources searched, the inclusion criteria, and the selected articles.

The first step involved scanning entries of the most common electronic databases of scientific literature, as well as the first 500 Google Scholar hits. The 4,856 resulting abstracts were analyzed, and all studies that did not violate the inclusion criteria were selected, resulting in a total of 325 articles, without duplicates and non-accessible full texts, available for further inspection. Two independent reviewers then reviewed these articles in depth with regard to the fulfillment of the inclusion criteria. This inspection yielded a total of 78 articles with 89 independent samples, including data of 5,973 participants. Most of the participants identified themselves as female (60%) and were university students (64%) with an overall mean age of 31.7 years.

All relevant data from these studies were summarized in a template to compute an effect size for each dependent variable examined. Based on this summary data, standardized mean differences between experimental groups exposed to robots varying in anthropomorphism were calculated. Most studies reported a comparison of means. However, the data sets were often incomplete, e.g., with no mention of means or standard deviations. Cohen's d was therefore chosen as a standard measure to describe the effect sizes. Note that Cohen's d represents an entire family of effect sizes, which makes it widely applicable for different study designs (e.g., Cohen's d_{av} for within study designs). In addition, it can be calculated from a wide range of statistical values received from different inferential statistical methods (e.g., ANOVAs or t-tests) (*52*). By using this measure as a standardized measure of effect sizes, we were able to compute a total of 294 effect sizes from the available data base. Different effect sizes derived from the same samples and similar outcome variables within a single study were averaged via the arithmetic mean. This was done so as not to overestimate those studies in comparison to others.

Overall, this resulted in a total of 187 effect sizes. The final set of effect sizes was then analyzed deductively by starting with the estimation of the overall mean effect of anthropomorphism on human-related outcomes via a random-effects model. The calculated mean effect size indicates the magnitude of the overall effect in terms of a standardized mean difference. If the 95% confidence interval does not include "zero", it can further be concluded that this mean difference indeed represents a statistically significant effect that can be expected to be replicated in further studies. To illustrate the effect size relative to its confidence interval, a forest plot was created. The square reflects the effect size; the size of the square shows the effect size weight with respect to the number of effect sizes included and confidence intervals are shown by the length of the whiskers (see Fig. 3 for illustration). In addition, the use of the random-effects model in this analysis also enabled us to assess the degree of heterogeneity of effect sizes. In contrast to random sampling errors as a cause of between-study differences, the heterogeneity estimates the true variation due to systematic differences in study design, sample, and measurements used (*53*, *54*). To estimate the level of heterogeneity, we used Q tests, which indicate whether or not a significant level of heterogeneity is present, and I^2, which represents the proportion of variance in the model that can be explained by unaccounted factors (54).

The second and third steps involved conducting a subset meta-analysis for each of the different superordinate outcome categories (i.e., perception, attitude, affect, behavior) and the respective subdimensions. Again, the analysis, based on random-effect models, allowed for assessing the mean effect sizes for different human-related outcomes and respective 95% confidence intervals, which were again illustrated via forest plots. In addition, we estimated the heterogeneity between the effects in different studies caused by hitherto unknown moderators.

Finally, a variety of moderator analyses were conducted, based on the set of possible moderators that had been identified a priori, i.e., the field of application, task relevance, morphology, and robot exposure. For the overall model, mixed-effect models were used for this purpose in order to include the moderators for diverting the directions or strength of the relationship between a predictor and an outcome (*53*, *55*). Moreover, we estimated the presence of heterogeneity via Q_M and the amount of heterogeneity via I^2 (in percent) accounted for by the different moderators. For the superordinate outcome categories, we abstained from using mixed-effect models, and limited our analysis to merely calculating the mean effect sizes and 95% confidence intervals in order to identify whether an effect was present at all. This somewhat constraint procedure was chosen because substantial heterogeneity in the data set can considerably reduce the statistical power of tests in mixed-effects models, which in turn would have increased the risk of failing to detect effects even if they were actually present (*56*).

In an additional analysis, the current data set was used to examine the degree of publication bias in the field of HRI. This was done because it has been suggested that unpublished results might systematically differ from published ones, especially because non-significant results may be submitted and published less frequently (*57*). Two different tools were used to detect such possible asymmetry between effects reported by published versus unpublished data, including a funnel plot to visually explore such bias and an Egger's regression test as an inferential statistical indicator. In the event of asymmetry, the two-sided trim-and-fill method was used to correct the data set for publication bias. This method is used to remove (trim) studies leading to asymmetry and replace the omitted studies (fill). It models the data as if effect sizes and standard errors were symmetrically distributed as they should be had all samples been unbiased estimators of the same mean value. As a result, the method generates an estimate of the number of missing studies and an adjusted effect size of a meta-analysis including the filled studies.

References

1. International Federation of Robotics, in *World Robotics 2020: Industrial Robots* (2020; https://ifr.org/img/worldrobotics/Executive_Summary_WR_2020_Industrial_Robots_1.pdf), pp. 13–16.
2. International Federation of Robotics, in *World Robotics 2020: Service Robots* (2020; https://ifr.org/img/worldrobotics/Executive_Summary_WR_2020_Service_Robots.pdf), pp. 11–12.
3. L. Onnasch, E. Roesler, A Taxonomy to Structure and Analyze Human–Robot Interaction. *Int. J. Soc. Robot.* **13**, 833–849 (2021).
4. T. Fong, I. Nourbakhsh, K. Dautenhahn, A survey of socially interactive robots. *Rob. Auton. Syst.* **42**, 143–166 (2003).
5. C. Breazeal, Toward sociable robots. *Rob. Auton. Syst.* **42**, 167–175 (2003).
6. V. Villani, F. Pini, F. Leali, C. Secchi, Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications. *Mechatronics*. **55**, 248–266 (2018).
7. E. Matheson, R. Minto, E. G. G. Zampieri, M. Faccio, G. Rosati, Human–robot collaboration in manufacturing applications: A review. *Robotics*. **8**, 100 (2019).
8. C. G. Atkeson, B. P. W. Babu, N. Banerjee, D. Berenson, C. P. Bove, X. Cui, M. DeDonato, R. Du, S. Feng, P. Franklin, M. Gennert, J. P. Graff, P. He, A. Jaeger, J. Kim, K. Knoedler, L. Li, C. Liu, X. Long, T. Padir, F. Polido, G. G. Tighe, X. Xinjilefu, in *IEEE-RAS International Conference on Humanoid Robots* (IEEE Computer Society, 2015), vols. 2015-December, pp. 623–630.
9. J. Luo, Y. Zhang, K. Hauser, H. A. Park, M. Paldhe, C. S. G. Lee, M. Grey, M. Stilman, J. H. Oh, J. Lee, I. Kim, P. Oh, in *Proceedings - IEEE International*

Conference on Robotics and Automation (Institute of Electrical and Electronics Engineers Inc., 2014), pp. 2792–2798.

10. G. Hoffman, C. Breazeal, in *AIAA 1st Intelligent Systems Technical Conference* (American Institute of Aeronautics and Astronautics, Reston, Virigina, 2004; http://arc.aiaa.org/doi/10.2514/6.2004-6434), p. 6434.
11. A. Clodic, E. Pacherie, R. Alami, R. Chatila, in *Sociality and Normativity for Robots: Studies in the Philosophy of Sociality*, R. Hakli, J. Seibt, Eds. (Springer, Cham, 2017; https://doi.org/10.1007/978-3-319-53133-5_8), pp. 159–177.
12. B. R. Duffy, Anthropomorphism and the social robot. *Rob. Auton. Syst.* **42**, 177–190 (2003).
13. A. Waytz, J. Cacioppo, N. Epley, Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspect. Psychol. Sci.* **5**, 219–232 (2010).
14. J. Złotowski, D. Proudfoot, K. Yogeeswaran, C. Bartneck, Anthropomorphism: Opportunities and challenges in human-robot Interaction. *Int. J. Soc. Robot.* **7**, 347–360 (2015).
15. E. Phillips, X. Zhao, D. Ullman, B. F. Malle, in *ACM/IEEE International Conference on Human-Robot Interaction* (IEEE Computer Society, 2018), pp. 105–113.
16. E. Roesler, J. I. Maier, L. Onnasch, The effect of anthropomorphism and failure comprehensibility on human-robot trust. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **64**, 107–111 (2020).
17. S. Stadler, A. Weiss, N. Mirnig, M. Tscheligi, in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (2013), pp. 231–232.
18. L. Onnasch, E. Roesler, Anthropomorphizing robots: The effect of framing in human-robot collaboration. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **63**, 1311–1315

(2019).

19. K. Darling, in *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence* (2017), pp. 173–188.
20. M. Mori, K. F. MacDorman, N. Kageki, The Uncanny Valley [From the Field]. *IEEE Robot. Autom. Mag.* **19**, 98–100 (2012).
21. K. S. Haring, D. Silvera-Tawil, K. Watanabe, M. Velonaki, in *Proceedings of the 2016 International Conference on Social Robotics: Lecture Notes in Computer Science*, A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs, H. He, Eds. (Springer, Cham, 2016), vol. 9979, pp. 392–401.
22. K. Darling, P. Nandy, C. Breazeal, in *Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (IEEE, 2015), pp. 770–775.
23. G. Hoffman, O. Zuckerman, G. Hirschberger, M. Luria, T. Shani-Sherman, in *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI ’15)* (2015), pp. 3–10.
24. T. Zhang, D. B. Kaber, B. Zhu, M. Swangnetr, P. Mosaly, L. Hodge, Service robot feature design effects on user perceptions and emotional responses. *Intell. Serv. Robot.* **3**, 73–88 (2010).
25. C. Bartneck, D. Kulić, E. Croft, S. Zoghbi, Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Robot.* **1**, 71–81 (2009).
26. T. Sanders, A. Kaplan, R. Koch, M. Schwartz, P. A. Hancock, The relationship between trust and use choice in human-robot interaction. *Hum. Factors*. **61**, 614–626 (2019).
27. P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. De Visser, R.

Parasuraman, A meta-analysis of factors affecting trust in human-robot interaction. *Hum. Factors*. **53**, 517–527 (2011).

28. G. Charalambous, S. Fletcher, P. Webb, The development of a scale to evaluate trust in industrial human-robot collaboration. *Int. J. Soc. Robot.* **8**, 193–209 (2016).

29. C. Bröhl, J. Nelles, C. Brandl, A. Mertens, C. M. Schlick, in *Proceedings of the HCI International 2016 – Posters' Extended Abstracts. Communications in Computer and Information Science*, C. Stephanidis, Ed. (Springer, 2016), vol. 617, pp. 97–103.

30. T. Nishida, Toward mutual dependency between empathy and technology. *AI Soc.* **28**, 277–287 (2013).

31. X. Dou, C.-F. Wu, K.-C. Lin, S. Gan, T.-M. Tseng, Effects of different types of social robot voices on affective evaluations in different application fields. *Int. J. Soc. Robot.* (2020), doi:10.1007/s12369-020-00654-9.

32. M. R. Fraune, S. Sherrin, S. Sabanovi, E. R. Smith, in *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)* (IEEE, New York, NY, USA, 2015), pp. 109–116.

33. S. L. Müller, S. Stiehm, S. Jeschke, A. Richert, in *Proceedings of the 2017 International Conference on Social Robotics: Lecture Notes in Computer Science* (Springer Verlag, 2017; https://doi.org/10.1007/978-3-319-70022-9_59), vol. 10652, pp. 597–606.

34. R. Reisenzein, Pleasure-Arousal Theory and the intensity of emotions. *J. Pers. Soc. Psychol.* **67**, 525–539 (1994).

35. J. A. Russell, A. Weiss, G. A. Mendelsohn, Affect grid: A single-item scale of pleasure and arousal. *J. Pers. Soc. Psychol.* **57**, 493–502 (1989).

36. S. Kuz, C. M. Schlick, in *Proceedings of the 19th Triennial Congress of the IEA* (Melbourne, 2015; https://www.researchgate.net/publication/282665386).

37. M. Salem, F. Eyssel, K. Rohlfing, S. Kopp, F. Joublin, To err is human(-like): Effects of robot gesture on perceived anthropomorphism and likability. *Int. J. Soc. Robot.* **5**, 313–323 (2013).

38. J. Fink, in *Proceedings of the 2012 International Conference on Social Robotics: Lecture Notes in Computer Science* (Springer, 2012), vol. 7621, pp. 199–208.

39. L. Onnasch, E. Roesler, A taxonomy to structure and analyze human–robot interaction. *Int. J. Soc. Robot.* (2020), doi:10.1007/s12369-020-00666-5.

40. International Organization for Standardization, ISO 8737:2012 Robots and robotic devices (2012), (available at https://www.iso.org/obp/ui/#iso:std:iso:8373:ed-2:v1:en).

41. M. Khoramshahi, A. Shukla, S. Raffard, B. G. Bardy, A. Billard, Role of gaze cues in interpersonal motor coordination: Towards higher affiliation in human-robot interaction. *PLoS One*. **11**, e0156874 (2016).

42. E. Roesler, J. I. Maier, L. Onnasch, in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (2020).

43. M. P. Mayer, S. Kuz, C. M. Schlick, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Springer, Berlin, Heidelberg, 2013; http://link.springer.com/10.1007/978-3-642-39182-8_11), vol. 8026 LNCS, pp. 93–100.

44. J. Li, The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *Int. J. Hum. Comput. Stud.* **77**, 23–37 (2015).

45. J. Wainer, D. J. Feil-Seifer, D. A. Shell, M. J. Matarić, in *Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2007)* (IEEE, 2007), pp. 872–877.

46. K. E. Schafer, T. Sanders, T. T. Kessler, M. Dunfee, T. Wild, P. A. Hancock, in

Proceedings of the 2015 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision (CogSIMA 2015) (IEEE, 2015), pp. 113–117.

47. C.-C. Ho, K. F. MacDorman, Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Comput. Human Behav.* **26**, 1508–1518 (2010).

48. A. Weiss, C. Bartneck, in *Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2015)* (IEEE, 2015), pp. 381–388.

49. P. Madden, L. Feingold, The Robots Are Here: At George Mason University, They Deliver Food To Students. *NPR* (2019), (available at https://www.npr.org/2019/04/07/710825996/the-robots-are-here-at-george-mason-university-they-deliver-food-to-students?t=1620380296102).

50. D. Moher, L. Shamseer, M. Clarke, D. Ghersi, A. Liberati, M. Petticrew, P. Shekelle, L. A. Stewart, P.-P. Group, Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* **4**, 1 (2015).

51. E. Roesler, L. Onnasch, D. Manzey, Same same, but different - A meta-analysis regaring the consequences of anthropomorphism in human robot interaction: PRISMA-P Protocol. *available at osf.io/egtk6* (2020).

52. D. Lakens, Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Front. Psychol.* **4**, 863 (2013).

53. S. Jain, S. K. Sharma, K. Jain, Meta-analysis of fixed, random and mixed effects models. *Int. J. Math. Eng. Manag. Sci.* **4**, 199–218 (2019).

54. H. Cooper, L. V. Hedges, J. C. Valentine, Eds., *The handbook of research synthesis and meta-analysis* (Russell Sage Foundation, New York, NY, USA, ed. 3, 2019).

55. W. Viechtbauer, Conducting meta-analyses in R with the Metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).

56. L. V. Hedges, T. D. Pigott, The power of statistical tests for moderators in meta-analysis. *Psychol. Methods*. **9**, 426–445 (2004).

57. K. Dickersin, in *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, H. R. Rothstein, A. J. Sutton, M. Borenstein, Eds. (John Wiley & Sons, Chichester, UK, 2006), pp. 9–33.

2.2 Publication 2: Why Context Matters: The Influence of Application Domain on Preferred Degree of Anthropomorphism and Gender Attribution in Human–Robot Interaction

Eileen Roesler, Naendrup-Poell, L., Dietrich Manzey & Linda Onnasch

Document type: Journal article | Published version (i.e., final, published version; also known as: Publisher PDF, Version of Record, Final Version)

International Journal of Social Robotics (2022) 14:1155–1166
https://doi.org/10.1007/s12369-021-00860-z

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

Eileen Roesler[1] · Lara Naendrup-Poell[1] · Dietrich Manzey[1] · Linda Onnasch[2]

Accepted: 23 December 2021 / Published online: 24 January 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., childminding). 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

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

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), 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 pretest 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 comparison 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 $\kappa_{Fleiss} = 0.74$. After discussing the diverging categories, in almost all cases an agreement was reached, resulting in a nearly perfect inter-rater reliability ($\kappa_{Fleiss} = 0.96$). 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 sociodemographic questions were filled in. The entire experiment lasted 15–20 min.

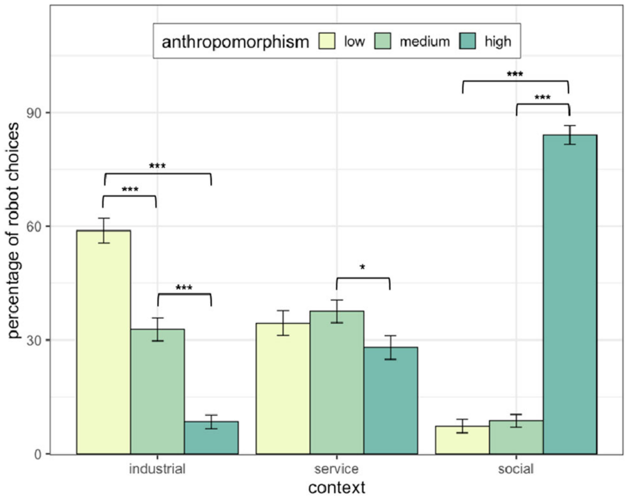

Fig. 3 Frequencies of robot selection (in case of anthropomorphism: low vs. medium vs. high) in percentage for the three context descriptions (industrial vs. service vs. social)

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

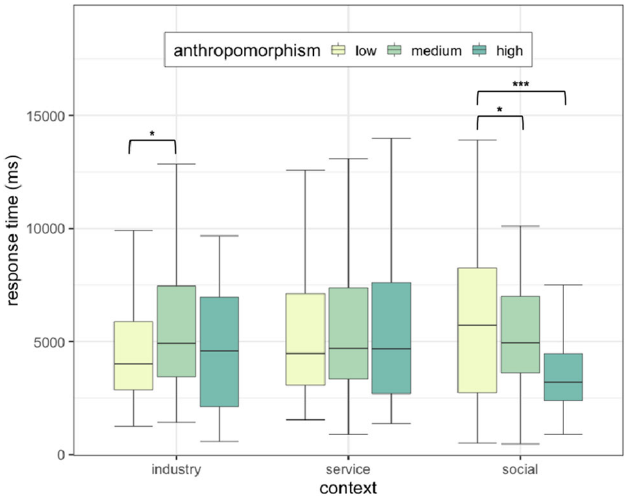

Fig. 4 Response times of robot selection (in case of anthropomorphism: low vs. medium vs. high) in ms for the three contexts (industrial vs. service vs. social)

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-

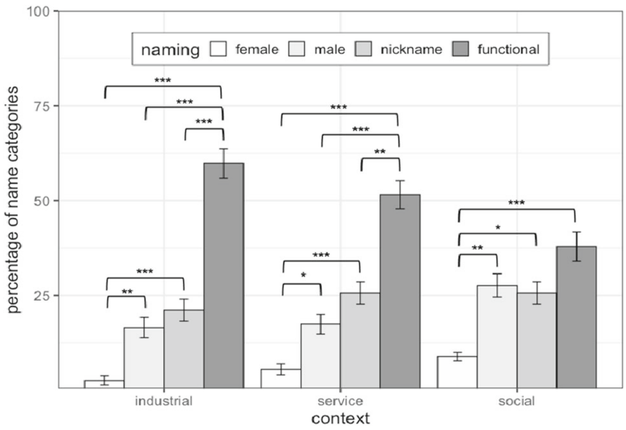

Fig. 5 Frequencies of chosen names (female vs. male vs. nickname vs. functional) in percentage for the three contexts (industrial vs. service vs. social)

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.

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
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/10.1177/1071181319631209
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: Robo-sexism 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

2.3 Publication 3: (Hu)man-like Robots — The Impact of Anthropomorphism and Language on Perceived Robot Gender

Roesler, E., Heuring, M., & Onnasch, L.

Document type: Journal article | Submitted version (i.e., Preprint after first round of revision) | Submitted to the International Journal of Social Robotics: Special Issue on GENDERING ROBOTS (GenR): Ongoing (Re)configurations of Gender in Robotics

2.4 Publication 4: Robots in Different Jobs — Associations of Anthropomorphism and Gender Stereotypes

Roesler, E., Heuring, M., & Onnasch, L.

Document type: Journal article | Submitted version (i.e., Preprint) | Submitted to Technology in Society

Title Page – Anonymized Submission

Title: Robots in Different Jobs — Associations of Anthropomorphism and Gender Stereotype

Authors: Eileen Roesler[a*], Maris Heuring[a], and Linda Onnasch[a]

[a]Department of Psychology and Ergonomics, Technische Universität Berlin, Marchstr. 12, 10587 Berlin, Germany
E-Mail-Addresses: eileen.roesler@tu-berlin.de, heuring@campus.tu-berlin.de, linda.onnasch@tu-berlin.de

[*] Corresponding Author

Declarations of interest: none.

Robots in Different Jobs — Associations of Anthropomorphism and Gender Stereotypes

ARTICLE INFO

Keywords:
human-robot interaction
anthropomorphism
application domain
task sociability
occupational gender stereotypes

ABSTRACT

Robots take over tasks in various application domains. In all domains, the tendency exists to equip robots with anthropomorphic appearance features. This trend is based on the assumption that anthropomorphism enhances the positive perception of robots as team partners. However, a growing number of studies have shown that this general assumption is not valid. Whereas anthropomorphic features seem to be desired in the social application domain, more technical robots are preferred in the industrial one. Currently, however, little is known about whether this effect is really based on the application domain or related to requirements of typical tasks in these domains. Therefore, this study investigated the influence of application domain and task sociability on the preferred degree of anthropomorphism. In an online study, participants received different descriptions of an industrial or social context with either high or low task sociability. Subsequently, they selected one robot out of differently anthropomorphic robots. In addition, we examined the influence of possible occupational gender stereotypes on gender-neutrally designed robots by asking participants to assign a name and a pronoun to their selected robot. The results indicated an overall preference for low anthropomorphic robots in the industrial domain, independently of the task sociability. In contrast, task sociability mattered for the social domain, where low anthropomorphic robots were preferred for tasks requiring low task sociability. In addition, the results showed a more functional than gendered association of robots. However, if robots were gendered, the assigned gender was male in the industrial domain, whereas no binary gender was assigned significantly more often in the social domain. In summary, the results showed that the preferred robot design depends on the domain the robot is deployed in and that even gender-neutral robots can elicit the occupational male-robot bias in the industrial domain.

1. Introduction

For most humans, the representation of robots is drawn more from science fiction than from actual interaction (Kriz, Ferro, Damera and Porter, 2010). However, fiction and reality differ considerably in regard to the design of robots. Since the first conception of robots that was introduced in the play *Rossum's Universal Robots*, fiction portrayed robots as replicas of humans (Reilly, 2019). In contrast, the robots that have most prominently conquered our private and working worlds are still rather technically designed. Whether private service robots in our households (e.g., for cleaning task), commercial service robots (e.g., for transporting goods) or industrial manufacturing robots (e.g., for pick and place tasks), all seem to have a technical rather than anthropomorphic appearance (Phillips, Zhao, Ullman and Malle, 2018). But there seems to be a shift of reality towards fiction as domain-overlapping more anthropomorphic features are implemented in robots (Millo, Gesualdo, Fraboni and Giusino, 2021). The idea behind decoratively adding anthropomorphic design is to increase the robots' perception as team partners and parts of our human society (Roesler, Manzey and Onnasch, 2021). This trend, though, does not seem to be desirable in all domains and for all kind of tasks (Goetz, Kiesler and Powers, 2003; Roesler, Naendrup-Poell, Manzey and Onnasch, 2022). An anthropomorphic design seems beneficial in the social domain but might be ineffective or detrimental in the industrial and service domain (Roesler et al., 2021). Recent studies have shown that anthropomorphic robot features like a face can decrease the perceived reliability of and trust in a robot in industrial settings (Roesler, Onnasch and Majer, 2020; Onnasch and Hildebrandt, 2021). Nonetheless, the assumption that anthropomorphism is a universal remedy to increase acceptance of robots as team partners often still remains.

For example, the company *Rethink Robotics* added facial features to its collaborative industrial robot *Sawyer* to increase acceptance. In particular, they state on their website: "The human workforce quickly accepts Sawyer, thanks to his friendly design." (Robotics, n.d.). This statement not only demonstrates that anthropomorphism is assumed to have positive effects on industrial robots, but also that this might be accompanied by a social categorization of robots in terms of gender. Besides this website, studies that describe *Sawyer* in an anthropomorphic manner also assign a male

ORCID(s):

pronoun to the robot (Kopp, Baumgartner and Kinkel, 2022). The fact that the robot is referenced with a masculine pronoun can have different reasons. First, *Sawyer* might be perceived as masculine rather than gender-neutral in its appearance (Perugia, Guidi, Bicchi and Parlangeli, 2022). Second, this might be related to the general association of masculinity and technology (Faulkner, 2001). Third, as *Sawyer* is an industrial robot, human occupational stereotypes might be transferred to it. Regardless of which of these reasons is the decisive one, it becomes evident that if we apply anthropomorphic features to robots, we are confronted with social issues of possibly gendered technologies. So if anthropomorphism is not clearly desired in all domains and for all tasks (Goetz et al., 2003; Roesler et al., 2022), we should consciously refrain from using this design. In order to make this conscious decision, we need to understand which factors of the application domain and the task are determining the desirability of anthropomorphic appearance from a human perspective.

1.1. Domain- and task-related preferences for anthropomorphic design

A major reason why anthropomorphic design is used extensively in all domains is the assumption that it triggers the activation of human-human interaction schemes (Haring, Watanabe and Mougenot, 2013; Fischer, 2022). Since humans are experts in interacting with other humans, anthropomorphic design is assumed to facilitate the intuitive interaction with novel technologies like robots. Although anthropomorphism can be implemented via various properties of the robot (e.g., movements or communication), one of the most salient properties is the visual appearance (Hegel, 2012). However, even if a positive effect of anthropomorphic appearance is often assumed, it is important that the appearance matches the requirements of the task the robot fulfills (Goetz et al., 2003; Złotowski, Khalil and Abdallah, 2020; Roesler et al., 2022). The broader frame of task requirements is set by the application domain in which a human-robot interaction (HRI) is situated (Onnasch and Roesler, 2021). That is, robots in the industrial domain need to be equipped with physical strength and precision to transport or manipulate work pieces. In comparison, robots in the social domain need to be equipped with communication capabilities to physically, emotionally and cognitively stimulate humans. Taking the requirements of both domains into account it seems reasonable that only the social domain broadly benefits from anthropomorphic robot features (i.e., in this case going beyond appearance by including also communication, movement, and context), whereas the results in the industrial domain are at least mixed (Roesler et al., 2021).

The usefulness of anthropomorphism in different application domains further seems to reflect people's preferences for robot design in the respective domains (Goetz et al., 2003; Roesler et al., 2022). The assumption that people expect robots to look appropriate for a given task context is referred to as the *matching hypothesis* (Goetz et al., 2003). The hypothesis is widely supported for both domains (i.e., industry, service, social) (Roesler et al., 2022) and specific tasks in the social domain (e.g., teacher or instructor) (Goetz et al., 2003; Złotowski et al., 2020). In line with this, in the current study we hypothesized that robots with a lower degree of anthropomorphism are preferred in the industrial domain. Vice versa, for the social domain, we assumed that robots with a higher degree of anthropomorphism are preferred.

Besides the general application domain, specific task characteristics also influence which robot appearance is preferred (Goetz et al., 2003; Złotowski et al., 2020). In particular, the task sociability seems to determine the perceived appropriateness of anthropomorphic appearance features (Goetz et al., 2003; Złotowski et al., 2020). Task sociability is described by the interaction time and frequency between interacting agents (human and / or technical) while performing a task (Li, Rau and Li, 2010), as well as the goal to affect people in social and cognitive terms (Breazeal, 2003).

Earlier studies on the impact of sociability focused rather on job sociability than task sociability. This, however, led to a confusion of sociability and other job characteristics (Goetz et al., 2003; Li et al., 2010; Złotowski et al., 2020). For example, Złotowski et al. (2020) operationalized low sociability with a security guard task, medium sociability with a tour guide as well as an entertainment task, and high sociability with a teaching task. These jobs have similarities and differences that go far beyond sociability. For example, comparable communicational requirements were present in one of the teaching tasks (i.e., requesting repetition of some English phrases) and one of the tour guide tasks (i.e., asking some questions), both namely representing the task goal of cognitive stimulation. Another teaching task, however, aimed for physical stimulation (i.e., requesting performance of physical exercises), while the entertaining task aimed for emotional stimulation (i.e., telling jokes). These examples illustrate that so far the fit of specific jobs with more or less anthropomorphic appearances has been studied (Goetz et al., 2003; Złotowski et al., 2020; Li et al., 2010) rather than the role of task sociability as a superordinate task feature. Thus, it is important to investigate the effect of task sociability on the preferred degree of anthropomorphism independently of other aspects (e.g., task specifications, autonomy or human role) (Onnasch and Roesler, 2019).

Moreover, for tasks in the industrial domain, task sociability was typically assumed to be low and for the social domain high (Breazeal, 2003). Newer generations of industrial robots challenge this dichotomy as they are used to closely and collaboratively interact with humans, too (Matheson, Minto, Zampieri, Faccio and Rosati, 2019). Tasks requiring high task sociability on the part of the robot are therefore not exclusively limited to one application domain anymore. However, little is currently known about the influence of task sociability on preferred robot design apart from the social and service domain (Goetz et al., 2003; Złotowski et al., 2020). This study, therefore, aimed to broaden the view to the industrial domain and to task sociability as a superordinate feature. Based on previous research (Goetz et al., 2003; Złotowski et al., 2020; Roesler et al., 2022), we assumed that for tasks requiring high task sociability, a higher degree of anthropomorphism is preferred over a lower one. Conversely, we hypothesized that for tasks requiring low task sociability, a lower degree of anthropomorphism is preferred over a higher one. In addition, this study aimed to investigate whether task sociability affects the preference for robot design differently in the industrial and social domains.

1.2. Occupational gender stereotypes in HRI

Both the industrial and social domain are associated with gender stereotypical occupational roles. The association of the industrial domain with masculinity and the social domain with femininity has various reasons. First, given that each domain is often saturated with persons of one gender (unbalanced female-to-male ratio), this gender is assumed to fit to the respective field better (Adachi, 2013; Lampousaki, 2010). Second, as different domains are assumed to have different requirements in regard to personal characteristics, gender-trait stereotypes (e.g., women are supposed to be nurturing) play an important role (Koenig, 2018). The assumption that a certain gender leads to more acceptance in a certain task is not only limited to human-human interaction. In HRI, it has also been postulated that endowing robots with gender-specific properties can facilitate the interaction (Eyssel and Hegel, 2012; Powers, Kramer, Lim, Kuo, lai Lee and Kiesler, 2005), especially if the occupational stereotype matches the manifested gender of the robot which performs the task. Accordingly, purposely gendered robots (via voice or appearance) are often perceived as more fitting for gender-stereotypical tasks (Tay, Jung and Park, 2014; Eyssel and Hegel, 2012; Chita-Tegmark, Lohani and Scheutz, 2019). For example, male-designed robots (via voice and name) were accepted more for a security task and female-designed robots for a healthcare task (Tay et al., 2014). However, robots that were used in these studies always provided obvious gender cues like gendered voices (Tay et al., 2014; Powers et al., 2005; Chita-Tegmark et al., 2019), hairstyles (Eyssel and Hegel, 2012), or lip color (Powers et al., 2005).

When we look at robots without obvious gendered cues, the research thins out. Therefore, empirical evidence for the unintended attribution of gender stereotypes to gender-neutral robots is rather scarce. One study addressing that question found a domain-overlapping male-robot bias (Roesler et al., 2022). However, this is only a single study and the study was conducted with a German sample. As already stated by other researchers (Reich-Stiebert and Eyssel, 2017), the word robot has a grammatical male gender in German. Even thought the neutral wording "the system" was used in earlier research (Roesler et al., 2022) the general association between robots and the grammatical gender of the word could have contributed to this bias. Therefore, the question remains open whether the transfer of occupational stereotypes or a general male-robot bias is also likely in an English sample. Based on the overall picture of research, we assumed that the perception of robots as male or female is triggered by the associated gender stereotype prevalent in the respective domain. In particular, we hypothesized that for the industrial domain, a male gender association with respect to robots exists. In contrast, we expected that a female gender ascription is present for robots in the social domain.

Furthermore, we aimed to broaden the view by taking the role of task sociability for possible gender stereotypes into account. One assumption that could be made from the presented research is that higher task sociability is associated with a female stereotype, as female-designed robots are considered to have a better fit for tasks to which high sociability is typically ascribed (Tay et al., 2014; Eyssel and Hegel, 2012). However, those results might have occurred based on the specific occupational stereotype (e.g., of the job, nurse or care taker as female-associated jobs) rather than its sociability. Moreover, and in contrast to a possible female-sociability bias, current research (Chita-Tegmark et al., 2019; Law, Chita-Tegmark and Scheutz, 2020) showed that male gendered robots were associated with higher emotional intelligence than female gendered robots. For this reason, it is difficult to make predictions about a specific gender effect for task sociability. We therefore only had the broad assumption, that overall more human gender associations (male, female, non-binary) compared to functional associations (neutral) are present for robots fulfilling more sociable tasks. In contrast, robots fulfilling less sociable tasks were expected to be associated in a functional rather than a gendered manner. On an exploratory level, we also investigate whether tasks with lower and higher task sociability are associated with a specific gender. With this, the current study aimed to investigate not only occupational, but also possible sociability-related stereotypes in HRI.

To investigate the interactional effects of the application domain and task sociability for the preferred degree of anthropomorphism and gender associations, an online study was conducted. In line with earlier research (Roesler et al., 2022), participants were presented with different vignettes describing a robot's task. Subsequently, the participants had to choose the robot that best fits to the described task in their opinion. The selection was limited to either a low or a medium anthropomorphic robot. No highly anthropomorphic robot stimuli were used, as most of them are perceived as gendered (Perugia et al., 2022). In the next step, the participants had to give a name and a pronoun to the robot. We extended the previously used naming technique by Roesler et al. (2022) with a pronoun technique, as we assumed that there could be differences between the assigned gender of the name and the gender implied by the pronoun. For instance, a robot could be given a gender-neutral or functional name, but still be referred to with the gendered pronouns "he" or "she". The pronoun technique is characterized by letting participants choose pronouns in a sentence referring to a presented agent. This technique has been used in psycholinguistic studies before (Flaherty, 2001; Moulton, Robinson and Elias, 1978), but has not yet been applied to the field of HRI. One advantage of this technique is that no coding post data collection is necessary because pronouns can be chosen from a fixed list. This might lead to a more objective assessment of gender than the naming technique, where especially gender-neutral vs. male/female categorization are often not clearly separable (Fleet and Atwater, 1997). Due to the introduction of chosen pronouns as a new measurement technique for assigned gender, an exploratory analysis was conducted to investigate the fit of both assigned names and pronouns.

2. Method

The experiment was preregistered and ethically evaluated via a checklist of the local ethics committee. The preregistration, ethical evaluation, and data can be accessed via the Open Science Framework (`https://osf.io/7fdyw/?view_only=9d50006dbb8b4c3d89bd783183ff7b96`).

2.1. Participants

An a priori power analysis for a *Pearson's* χ^2 *test* with an alpha level of $\alpha = .05$, three degrees of freedom, and a medium effect size of $w = .3$ showed that a total sample size of 120 participants was needed to achieve a power of .80. Participants were recruited through the platform *Prolific*. Only English-speaking individuals between 18 and 45 could take part. Of 130 complete cases, three were excluded due to invalid scores. Two more participants were excluded, as they entered "??" as robot names which could not be assigned to the defined naming categories. This left a sample of $N = 125$ with 63 participants in the low task sociability condition and 62 in the high task sociability condition. The mean age was 28.7 years, with a *SD* of 6.9 years. More women (77) than men (47) and non-binary individuals (1) took part in the study. A majority of the sample (58.4%) had a work or study background in the social domain, compared to 16.8% in the industrial domain, and to 24.8% in a different domain. Most participants were either university students (32%) or employees (36.8%). Each participant received a monetary reimbursement (£1.37) for completing the 10-minute study.

2.2. Task and materials

2.2.1. Scenario descriptions

Application domain and task sociability were manipulated using a combination of task and domain descriptions that introduced a scenario for the participant. This was necessary as individuals could not be expected to have similar ideas about concepts like application domain or task sociability. As the study design was comparable to (Roesler et al., 2022), we decided to use an altered version of the domain description from that study and further added the task descriptions. Each description initially outlined the domain and continued with a specific depiction of a task with a robot and a human as collaboratively working agents. The descriptions of both application contexts had to reflect the fields authentically: Inspiration was taken from literature concerning the industrial (Hägele, Nilsson and Pires, 2008; Stadler, Weiss, Mirnig and Tscheligi, 2013) and the social context (Van der Loos and Reinkensmeyer, 2008; Breazeal, Takanishi and Kobayashi, 2008). Moreover, the tasks had to reflect actual tasks that can (possibly) occur in each context and reflect different levels of sociability. In this study, task sociability was operationalized as "interaction time and frequency" in accordance with Li et al. (2010). In the latter study, the authors categorized jobs with frequent and intense interaction as jobs with high task sociability (most commonly teachers/instructors). In jobs with low task

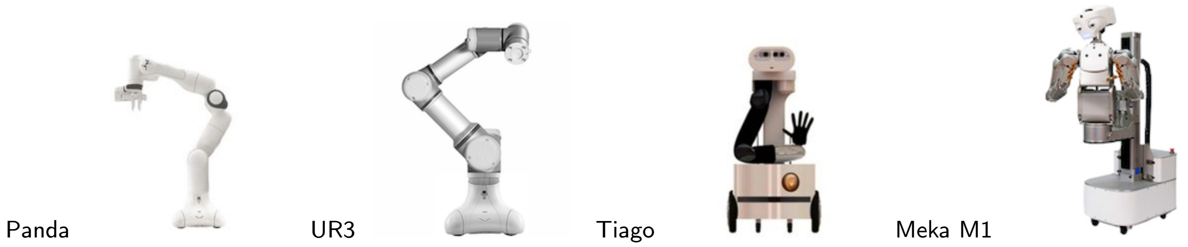

Figure 1: Stimuli retrieved from *ABOT* database for low (left) and medium (right) anthropomorphic robots (Phillips et al., 2018)

sociability (e.g., servant), interaction was shorter and less frequent. In the end, a pick and place task was chosen for the low task sociability and a teaching task was chosen for the high task sociability condition. All four texts were proofread by a British first-language English speaker and can be assessed via A. Appendix—Vignettes.

2.2.2. Robotic stimuli

Four images of real world robots (i.e., *Panda, UR3, Tiago,* and *Meka M1*, see Fig. 1) were presented to the participants over the course of the experiment. Those images were chosen based on a previous study by the authors using images from the *Anthropomorphic roBOT (ABOT)* database, as they are all perceived as gender-neutral (Perugia et al., 2022). Each participant had to choose an either a low or a medium anthropomorphic robot for each application domain. The scores within each category were comparable, whereas the scores between the low (M=7.28) and medium (M=24.37) level of anthropomorphism differed considerably. The order of presented robots was randomized for the selection.

2.3. Study design

A 2 x 2 mixed design was chosen, with the within factor application domain (industrial vs. social) and the between factor task sociability (low vs. high). Group assignment (low vs. high task sociability) and domain description order (industrial vs. social) were randomized.

2.4. Dependent measures

As a control variable, a 5-item short version of the Individual Differences in Anthropomorphism Questionnaire (IDAQ) (Waytz, Cacioppo and Epley, 2010), which only included questions referring to technological devices, was used to check for group differences between the low task sociability and the high task sociability sample in their tendency to anthropomorphize (Waytz et al., 2010). The preferred anthropomorphism (low vs. medium) was assessed via a forced picture choice between two out of the four robot pictures (randomized). For the analysis, the choice frequency with which the two different degrees of anthropomorphism were chosen across the conditions was computed. For measuring assigned gender, two techniques were applied – one assessing given names (i.e., free text) and one assigned pronouns (i.e., drop down menu). Given names were coded into four categories by three raters working in our research group. The categories were *female, male, functional*, and *gender-neutral* names. As an assessment of the inter-rater reliability of the three raters, *Fleiss' Kappa* was computed. After the first iteration, inter-rater reliability was substantial for names given in the industrial domain (K_{Fleiss} = .69) and already high in the social domain (K_{Fleiss} = .82). For the second iteration, every case of disagreement was discussed. High inter-rater reliability was reached in both the industrial (K_{Fleiss} = .81) and the social domain (K_{Fleiss} = .96). In case of discordance, the rating two out of three raters agreed upon was applied as the final rating. From these categories, frequencies of given names were computed. The pronoun options were *she, he, it*, and *they*. Frequencies of given pronouns were computed from these categories. In addition, the number of trials in which the gender of the assigned pronoun and name matched was computed.

2.5. Procedure

The online study was carried out using the platform *SoSciSurvey*. After consenting to the study, participants were presented with the instruction for the selection task. Every participant was assigned to one task sociability condition (low vs. high). Subsequently, participants were shown the domain description for either the industrial or the social application domain first. On the next page, they chose one out of two robots (low vs. medium anthropomorphic),

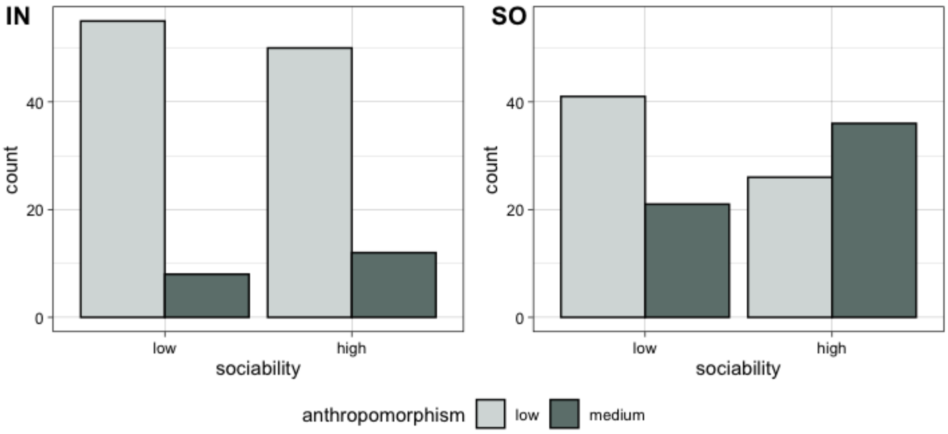

Figure 2: Robot selection (in the case of anthropomorphism: low vs. medium) in absolute numbers for both levels of sociability (low vs. high), separately for the industrial (IN) and social (SO) application domain.

which they found appropriate for the task at hand. For this robot, participants were asked to fill in the blanks in a sentence, choosing a name and a pronoun. Following that, the selection task was repeated with the domain description for the other domain. The experiment concluded with a short version of the IDAQ and a small set of sociodemographic questions (gender, age, field of work, employment, study program if applicable). Data collection took place in summer 2021.

3. Results

3.1. Control variable

To control for individual differences in the tendency to anthropomorphize, a shortened version of the IDAQ was used, which showed a good internal consistency ($\alpha = .81$). An independent t-test, comparing the mean IDAQ sum scores of the low ($M = 11.6$, $SD = 8.5$) and high sociability condition ($M = 11.5$, $SD = 7.7$) revealed no significant difference ($t(123) = 0.05$, $p = .961$).

3.2. Preferred degree of anthropomorphism

Figure 2 shows the preferred level of anthropomorphism for the different levels of sociability within the respective application domain. To investigate differences in the preferred degree of anthropomorphism between the application domains and levels of sociability multiple χ^2 tests were conducted (see Table 1).

Table 1
Comparisons of robot selection in the case of anthropomorphism separately for the industrial and social application domain and the respective levels of sociability (overall, low, and high). Asterisks indicate significance of p-values: * = $p < .05$, ** = $p < .01$, and *** = $p < .001$.

	Industrial Domain	Social Domain
Overall	$\chi^2(1, N = 125) = 57.8, p < .001^{***}$	$\chi^2(1, N = 125) = 1.0, p = .325$
Low Sociability	$\chi^2(1, N = 63) = 35.1, p < .001^{***}$	$\chi^2(1, N = 63) = 7.0, p = .008^{**}$
High Sociability	$\chi^2(1, N = 62) = 23.3, p < .001^{***}$	$\chi^2(1, N = 62) = 1.6, p = .204$

In the industrial domain, significantly more participants – 84% –preferred a low anthropomorphic robot over a medium anthropomorphic one. This pattern was also apparent in both sociability conditions, as 87.3% in the low sociability condition, and 80.6% in the high sociability condition preferred low anthropomorphic robots over medium anthropomorphic ones.

The analysis of the social application domain revealed overall no significant differences in choice frequencies. Participants did not choose significantly more robots with a medium degree of anthropomorphism (45.6%) compared

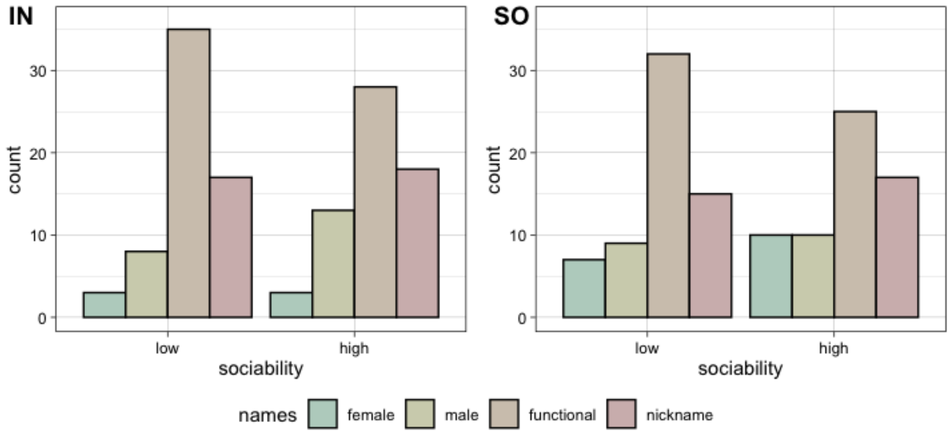

Figure 3: Chosen names (female vs. male vs. functional vs. nickname) in absolute numbers for both levels of sociability (low vs. high), separately for the industrial (IN) and social (SO) application domain.

to robots with a low degree of anthropomorphism. However, in the low sociability condition, a significant difference was revealed. As expected, more participants selected low anthropomorphic robots (66.7%) than medium anthropomorphic ones. Surprisingly, this was not mirrored in the high sociability condition, as participants did not choose significantly more often medium anthropomorphic robots (58.1%) compared to low anthropomorphic ones.

3.3. Names and pronouns

The coded gender of the names (see Figure 3) and pronouns (see Figure 4) assigned to the selected robot were analyzed in three steps. First, to determine whether occupational gender stereotypes exist in the domains and to exploratively investigate the role of sociability, the frequencies of assigned gender (i.e., female/she, male/he, functional/it, nicknames/they) were analyzed via multiple χ^2 tests for names (see Table 2) and pronouns (see Table. 3). For all analyses in which significant differences on the four-category dependent variables occurred, Bonferroni-corrected post-hoc pairwise comparisons were conducted. Second, to investigate whether sociability, and on an exploratory basis application domain, are associated with a human ascription, all human gender categories (i.e., female/she, male/he, nicknames/they) were combined and compared to functional associations. Again, multiple χ^2 tests were conducted for both assigned names (see Table 4) and pronouns (see Table 5). Third, due to the introduction of chosen pronouns as a new measurement technique for assigned gender, an exploratory correlation analysis was performed to investigate the fit of both assigned names and pronouns.

3.3.1. Gender ascription

Table 2
Comparisons of robot name assignment separately for the industrial and social application domain and the respective levels of sociability. Asterisks indicate significance of p-values: * = $p < .05$, ** = $p < .01$, and *** = $p < .001$.

	Industrial Domain	Social Domain
Overall	$\chi^2(3, N = 125) = 56.5, p < .001^{***}$	$\chi^2(3, N = 125) = 32.5, p < .001^{***}$
Low Sociability	$\chi^2(3, N = 63) = 37.8, p < .001^{***}$	$\chi^2(3, N = 63) = 24.6, p < .001^{***}$
High Sociability	$\chi^2(3, N = 62) = 21.0, p < .001^{***}$	$\chi^2(3, N = 62) = 9.9, p = .020^{**}$

Overall, the analyses of assigned names in the industrial domain revealed that female names (4.8%) were chosen significantly less often than any other type of name (all $ps < .024$). Conversely, functional names (50.4%) were chosen significantly more often than any other type of name (all $ps < .024$). For industrial tasks with low sociability, significantly more functional (55.6%) compared to both female (4.8%), and male (12.7%) names were assigned (all $ps < .001$). In addition, significantly more neutral (27.0%, $p < .001$) than female names were assigned. For tasks

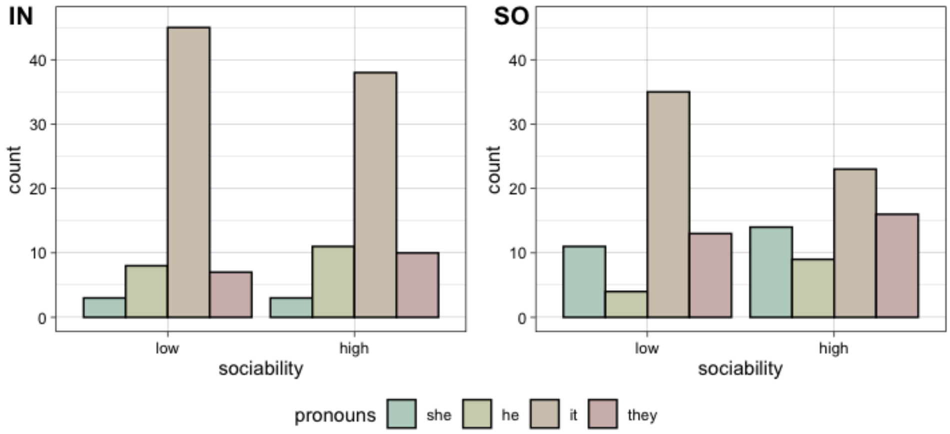

Figure 4: Chosen pronouns (she vs. he vs. it vs. they) in absolute numbers for both levels of sociability (low vs. high), separately for the industrial (IN) and social (SO) application domain.

requiring high sociability, again, only 4.8% of participants chose a female name. This was significantly less often than the assignment of functional (45.2%, $p < .001$) as well as neutral (29.0%, $p = .006$) names. In addition, participants descriptively tended to assign less often female than male names (21.0%, $p = .075$).

For the social domain, overall, functional names were chosen 45.6% of the time, significantly more often than all other types of name (all $ps < .049$) The analysis of the low sociability condition showed that functional names (50.8%) were chosen significantly more often than male (14.3%, $p = .002$) and female (11.1%, $p < .001$) names. The post-hoc comparisons of the high sociability condition showed no significant differences. On a descriptive level, female and male names (both 16.1%) were chosen less often than functional ones (40.3%, $ps = .067$).

Table 3
Comparisons of robot pronoun assignment separately for the industrial and social application domain and the respective levels of sociability. Asterisks indicate significance of p-values: * = $p < .05$, ** = $p < .01$, and *** = $p < .001$.

	Industrial Domain	Social Domain
Overall	$\chi^2(3, N = 125) = 117.4, p < .001^{***}$	$\chi^2(3, N = 125) = 35.0, p < .001^{***}$
Low Sociability	$\chi^2(3, N = 63) = 73.3, p < .001^{***}$	$\chi^2(3, N = 63) = 34.2, p < .001^{***}$
High Sociability	$\chi^2(3, N = 62) = 46.0, p < .001^{***}$	$\chi^2(3, N = 62) = 6.5, p = .090$

For assigned pronouns, the analyses of the industrial domain revealed that overall functional pronouns (66.4%) were assigned significantly more often than all other types of pronouns (all $ps < .001$). In addition, there was a trend of more male than female pronouns assigned in this application domain, that just failed to reach the conventional level of significance ($p = .056$). Moreover, functional pronouns were assigned significantly more often than any other type of pronoun (all $ps < .001$) for tasks with both low (71.4%) and high sociability (61.3%).

In line with the results of the industrial domain, functional pronouns (46.4%) were assigned significantly more often than any other type of pronoun (all $ps < .012$) in the social domain. Also, the analysis of tasks requiring low sociability revealed that functional pronouns (55.6%) were used significantly more often than any other type of pronoun (all $ps < .010$). In contrast to this, no significant differences were revealed for tasks requiring high sociability.

3.3.2. Human ascription

For the industrial domain, no significant differences in the choice frequency of human and functional names were revealed. This was also the case for both the low and high sociability condition. In line with this, the analysis of the social application domain revealed no significant differences between human and functional names. Again, no differences were found for the low and high sociability condition.

Table 4
Comparisons of robot name assignment in regard to humanness separately for the industrial and social application domain and the respective levels of sociability. Asterisks indicate significance of p-values: * = $p < .05$, ** = $p < .01$, and *** = $p < .001$.

	Industrial Domain	Social Domain
Overall	$\chi^2(1, N = 125) = 0.1, p = .929$	$\chi^2(1, N = 125) = 1.0, p = .325$
Low Sociability	$\chi^2(1, N = 63) = 0.8, p = .378$	$\chi^2(1, N = 63) = 0.1, p = .900$
High Sociability	$(\chi^2(1, N = 62) = 0.6, p = .446$	$\chi^2(1, N = 62) = 2.3, p = .127$

Table 5
Comparisons of robot pronoun assignment in regard to humanness separately for the industrial and social application domain and the respective levels of sociability. Asterisks indicate significance of p-values: * = $p < .05$, ** = $p < .01$, and *** = $p < .001$.

	Industrial Domain	Social Domain
Overall	$\chi^2(1, N = 125) = 13.4, p < .001^{***}$	$\chi^2(1, N = 125) = 0.7, p = .421$
Low Sociability	$\chi^2(1, N = 63) = 11.6, p < .001^{***}$	$\chi^2(1, N = 63) = 0.8, p = .378$
High Sociability	$\chi^2(1, N = 62) = 3.2, p = .075$	$\chi^2(1, N = 62) = 4.1, p = .042^{*}$

In contrast to the results of the naming task, a significant difference between functional and human pronouns was revealed in the industrial domain. That is, functional pronouns (66.4%) were assigned significantly more often than human pronouns. Further analysis revealed that this is the case for tasks requiring low task sociability, and only descriptively for tasks requiring high task sociability. In the low task sociability condition, significantly more participants assigned functional (71.4%) compared to human pronouns. This was only descriptively shown in the high task sociability condition, as functional pronouns (61.3%) were assigned slightly more than human pronouns.

Overall, the analysis of the social domain revealed no significant difference between functional and human pronouns. In line with this, the analysis of the low sociability condition showed no significant difference either. In contrast, participants assigned significantly more human (62.9%) than functional pronouns in the high sociability condition.

3.3.3. Name-pronoun association

For the industrial domain, a moderate association between names and pronouns was revealed (V=.456). More precisely, 58.4% of all cases had a perfect match of name and pronoun. Closer inspections of the mismatching cases showed that most often participants mixed functional and gender-neutral pronouns and nouns (20.8%). Interestingly, gender-neutral nicknames were combined more commonly with male pronouns (6.4%) than female pronouns (0.8%). For the social domain, a moderate association between names and pronouns was found, too (V=.427). Again, in most of the cases (52.8%), the chosen name and assigned pronoun matched. In 20% of the cases, participants combined functional pronouns/names with gender-neutral pronouns/names. In contrast to the industrial domain, gender-neutral names were combined more often with female (5.6%) than male (2.4%) pronouns in the social domain. In both application domains, female and male names were never assigned the opposite gendered pronoun.

4. Discussion

As robots increasingly play a role in our personal and professional lives, it is important to understand which robotic appearance is preferred in different domains and for different tasks. Previous research primarily focused on the role of different application domains (Roesler et al., 2022) or rather specific tasks (Goetz et al., 2003; Złotowski et al., 2020; Li et al., 2010). The current experiment aimed to broaden the view to an underlying property of different tasks – the sociability. In particular, the relationship between this task characteristic and the application domain in regard to the preferred robot appearance was addressed in our research.

The general assumption was based on the *matching hypothesis* (Goetz et al., 2003), which suggests that a preference for more anthropomorphic robots would be apparent in the social domain as well as for tasks with high task sociability. Vice versa, less anthropomorphic robots were assumed to be preferred in the industrial domain as well as for tasks with low task sociability. However, the results are only partially in line with this hypothesis (Goetz et al., 2003) and earlier research (Roesler et al., 2022; Złotowski et al., 2020; Li et al., 2010). First and as hypothesized, the results indicated a significant preference for robots with a lower degree of anthropomorphism in the industrial domain. Unexpectedly, this was the case independently of the tasks' sociability. One reason for the non-significant effect of task sociability in the industrial domain might be related to our chosen stimuli. Robotic arms like *Panda* and *UR3* are strongly associated with the industrial domain (Hägele et al., 2008). Therefore, existing knowledge and/or experience of what industrial robots actually look like might have been more dominant in the decision than the sociability of the task. Second, and contrary to earlier research conducted by Roesler et al. (2022), robots with a higher degree of anthropomorphism were not generally preferred in the social domain. In contrast to the industrial domain, however, sociability seemed to be the decisive factor in this case. Robots with lower degrees of anthropomorphism were significantly preferred for tasks with low sociability. Surprisingly, this effect was not significantly mirrored for tasks with high sociability, which is in contrast to earlier research (Roesler et al., 2022; Złotowski et al., 2020; Goetz et al., 2003). Again, this might be related to the stimuli used in the current experiment. A recent study (Kim, Bruce, Brown, de Visser and Phillips, 2020) investigating the uncanny valley, based on all stimuli of the ABOT database (Phillips et al., 2018), found an early uncanny valley for robots that rather moderately resemble human appearance. Both medium anthropomorphic robots used in this experiment are in the range of this early uncanny valley. In addition, the gender-neutrality of our stimuli might have further contributed to the assumed negative effect, as ambiguous gender cues can lead to increased eeriness (Paetzel, Peters, Nyström and Castellano, 2016). The possibility that the medium anthropomorphic robots might have induced a certain degree of eeriness is, however, limited to values far away from the original uncanny valley (Kim et al., 2020).

Despite the methodological limitations in regard to the stimuli and the online setting of the study, the results further strengthen the current body of research, which illustrates that the desirability of anthropomorphic appearance is highly context-sensitive (Goetz et al., 2003; Roesler et al., 2021, 2022). In detail, decorative anthropomorphic features without task-relevance (e.g., no increased predictability of robot behavior) should be avoided in industrial HRI. Especially, as this study found no preference for more anthropomorphic robots and other studies already illustrated potential negative effects in regard to trust, perceived reliability, and attention allocation (Roesler et al., 2020; Onnasch and Hildebrandt, 2021). Additionally, the results contribute to a more nuanced view on the role of anthropomorphism in the social domain. The overall positive effect of anthropomorphism (Roesler et al., 2021) and the preference for anthropomorphic robots (Roesler et al., 2022) seem to be related rather to the required task sociability (Goetz et al., 2003; Złotowski et al., 2020) than to the application domain in general.

However, the application domain in general might lead to the activation of occupational stereotypes in HRI (Tay et al., 2014; Roesler et al., 2022). This assumption was partially supported by the results of the current study. Most prominently and in line with earlier research (Roesler et al., 2022), functional and gender-neutral names and pronouns were assigned to robots in both domains. This is not surprising considering that the robotic stimuli were extensively pretested and compared to a recent data set (Perugia et al., 2022) to ensure that they are perceived as gender-neutral. When we take a closer look at the conditions under which the chosen robots were assigned a binary gender, an interesting pattern emerges. In contrast to other research, the results did confirm neither a general male-robot bias (Ladwig and Ferstl, 2018; Roesler et al., 2022; Perugia et al., 2022) nor a one-on-one transfer of occupational stereotypes (Tay et al., 2014). The transfer of occupational stereotypes seems to be present in the industrial domain, as the assigned names significantly showed and pronouns tended in this direction. So if robots are gendered in the industrial domain, they are assigned more male than female names and pronouns. In addition, this seems to be mostly apparent for tasks requiring higher task sociability. It can be assumed that in this domain both the association of technology (like robots) and masculinity (Faulkner, 2001) and the occupational stereotype of men are present. The concurrent presence of the male-technology bias and the occupational stereotype of women in the social domain might have led to no significant differences in the social domain between male and female associations. Neither the assigned names nor pronouns showed a significant dominance for a binary gender association. This was also the case in both sociability conditions.

Additionally, we assumed that higher task sociability would lead to more human gender ascription, as those tasks seem to be associated with human likeness (Goetz et al., 2003). However, this could not be confirmed. Higher task sociability did not lead to the ascription of significantly more human names. This was the case for both the industrial and the social domain. However, the analysis of the assigned pronouns showed that in the industrial domain more functional than human pronouns were assigned. This was significantly shown for tasks requiring low sociability, and descriptively for tasks with high sociability. In line with a preference for technical robots in this domain, more functional than human pronouns were used, seemingly independent of the sociability. In contrast, and in line with the results in regard to anthropomorphic appearance, sociability mattered for the social domain. More human pronouns were assigned for tasks requiring high task sociability, but not vice versa, more functional pronouns for tasks requiring low task sociability. Whereas the naming technique did not reveal any significant results or trends, the pronoun technique revealed differences in the ascription of human names. This further indicates that the ascription of names or nicknames does not need to perfectly go along with the usage of human pronouns (Sung, Guo, Grinter and Christensen, 2007).

Overall, the combination of names and pronouns can be considered beneficial, as a combination can strengthen the generalizability of results and each method can partially compensate for the shortcomings of the other. First, the categories of names and pronouns directly corresponded by more than 50% and binary gender ascription were never mixed (e.g., male pronoun and female name), indicating a high precision for detecting binary gender attributions. Second, pronouns offer a more objective assessment (Fleet and Atwater, 1997) than the naming technique, as no additional coding is required. Third, compared to the pronoun technique, naming might be less prone to response biases like social desirability due to the free text format. This might also be a reason why pronouns just indicated a non-significant trend towards more male pronouns in the industrial domain, whereas the naming technique revealed a significant effect. For these reasons, the combination of both methods might be fruitful for future research to investigate gender biases in HRI.

Taken together, the results illustrated that name and pronoun ascription, as well as preferred degree of anthropomorphism, depend on the job the robot is assumed to fulfill. Whereas a preference for low anthropomorphic robots and functional pronouns exists in the industrial domain, sociability matters in the social domain. In this domain, less anthropomorphic robots are preferred for tasks with low sociability, and more human pronoun ascription is present for tasks with high sociability. Moreover, a male-robot bias was revealed in the industrial domain, whereas no occupational bias was found in the social domain. However, it is important to have the limitations in mind to evaluate the validity of these results.

This study was an online vignette study, which only incorporated pictures of real robots (Phillips et al., 2018). Even though depictions of robots are generally useful to investigate the effects of anthropomorphism on a subjective level (Roesler et al., 2021, 2022), an extensive presentation of the robot beyond the appearance is not possible. Given the fact that anthropomorphism can be implemented additionally via movements, communication, and context (Onnasch and Roesler, 2021), the current results only allow assumptions to be made about anthropomorphic appearance features. In particular, robotic voices are an important feature that genders robots (Bryant, Borenstein and Howard, 2020; Tay et al., 2014), and the tendency that non-gendered robot voices are perceived as male (Chita-Tegmark et al., 2019) opens up multiple avenues for future research. Besides anthropomorphism, a plethora of other factors, like likability or eeriness (Złotowski et al., 2020), influence the preferences for different robot appearances. Since no data is yet available in the ABOT database about the mentioned factors (Phillips et al., 2018), it cannot be ruled out that some robot stimuli were preferred due to other aspects than their anthropomorphism. Lastly, no highly anthropomorphic stimuli were used, as most of them were perceived as gendered (Perugia et al., 2022). Therefore, it remains unclear if highly anthropomorphic robots would be preferred under some circumstances, even though they might be gendered.

For low and medium anthropomorphic robots, the results showed that a gender-neutral appearance elicits primarily functional and gender-neutral associations. Even though a mismatch of robot gender and occupational stereotype might be beneficial in some contexts (Reich-Stiebert and Eyssel, 2017; Powers et al., 2005), a gender-neutral appearance offers the opportunity to reduce or even avoid gender stereotypes in HRI (Darling, 2015). Also, beyond the problematic effect of enhancing gender stereotypes, the question remains whether anthropomorphism is always beneficial. The current study not only supports the findings that the effectiveness of and preference for anthropomorphism is domain and task specific (Goetz et al., 2003; Roesler et al., 2021, 2022; Złotowski et al., 2020), but also extends the body of research by the interrelation of domain and task sociability. For the industrial domain, robots with lower degrees of anthropomorphism are preferred independently of the sociability of the task. However, sociability matters for the social domain, as low anthropomorphic robots are preferred in tasks requiring low sociability. In conclusion, our results illustrate the importance of context factors in HRI and the advantages of gender-neutral robot design.

A. Appendix - Vignettes

Industrial Domain x Low Task Sociability: Soon, highly automated technical support is to be introduced in the industrial work context. Possible places could be factories or workshops, where metals or polymers are processed. The robotic assistance helps humans to assemble products at the assembly line. 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. From there, the human takes the parts and puts them together.

Social Domain x Low task Sociability: For work in the social field, a highly automated assistance system will soon be introduced. Possible places could be hospitals or senior citizens' residences. Work centers care for humans in need. The system supports the nursing staff of an institution with handing out medication. This robotic assistant prepares the right preparation in the right dosage, i.e., searches the required drug, defines the quantity and places it in a designated area. From there, the human takes the medication and gives it to the people that need it in the institution.

Industrial Domain x High Task Sociability: Soon, highly automated technical support is to be introduced in the industrial work context. Possible places could be factories or workshops, where metals or polymers are processed. The robotic assistance helps the human to assemble products at the assembly line. The robotic assistance helps the human at the assembly line, when a new production phase is introduced. The system gives the human instructions for carrying out the tasks and demonstrates them at the assembly line. It also gives advice on optimal ergonomic execution to support the human.

Social Domain x High Task Sociability: For work in the social field, a highly automated assistance system will soon be introduced. Possible places could be hospitals or senior citizens' residences. Work centers care for humans in need. The system supports the nursing staff of an institution with care work, as it helps humans to learn and execute physical exercise to maintain fitness. The robotic assistant gives the human instructions for carrying out the tasks and demonstrates them itself. It also gives advice on optimal ergonomic execution to support the human.

CRediT authorship contribution statement

: Conceptualization of this study, Methodology, Validation, Formal analysis, Visualisation, Writing - Original Draft, Writing - Review & Editing. **:** Software, Investigation, Writing - Review & Editing. **:** Writing - Review & Editing, Supervision.

References

Adachi, T., 2013. Occupational gender stereotypes: Is the ratio of women to men a powerful determinant? Psychological Reports 112, 640–650. doi:10.2466/17.07.pr0.112.2.640-650.

Breazeal, C., 2003. Toward sociable robots. Robotics and Autonomous Systems 42, 167–175. doi:10.1016/S0921-8890(02)00373-1.

Breazeal, C., Takanishi, A., Kobayashi, T., 2008. Social Robots that Interact with People, in: Siciliano, B., Khatib, O. (Eds.), Springer Handbook of Robotics. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1349–1369. doi:10.1007/978-3-540-30301-5_59.

Bryant, D., Borenstein, J., Howard, A., 2020. Why should we gender?, in: Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, ACM. doi:10.1145/3319502.3374778.

Chita-Tegmark, M., Lohani, M., Scheutz, M., 2019. Gender effects in perceptions of robots and humans with varying emotional intelligence, in: 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE. doi:10.1109/hri.2019.8673222.

Darling, K., 2015. 'who's johnny?' anthropomorphic framing in human-robot interaction, integration, and policy. SSRN Electronic Journal doi:10.2139/ssrn.2588669.

Eyssel, F., Hegel, F., 2012. (s)he's got the look: Gender stereotyping of robots. Journal of Applied Social Psychology 42, 2213–2230. doi:10.1111/j.1559-1816.2012.00937.x.

Faulkner, W., 2001. The technology question in feminism. Women's Studies International Forum 24, 79–95. doi:10.1016/s0277-5395(00)00166-7.

Fischer, K., 2022. Tracking anthropomorphizing behavior in human-robot interaction. J. Hum.-Robot Interact. 11. doi:10.1145/3442677.

Flaherty, M., 2001. How a language gender system creeps into perception. Journal of Cross-Cultural Psychology 32, 18–31. doi:10.1177/0022022101032001005.

Fleet, D.D.V., Atwater, L., 1997. Sex Roles 37, 111–123. doi:10.1023/a:1025696905342.

Goetz, J., Kiesler, S., Powers, A., 2003. Matching robot appearance and behavior to tasks to improve human-robot cooperation, IEEE. pp. 55–60.

Haring, K.S., Watanabe, K., Mougenot, C., 2013. The influence of robot appearance on assessment, in: Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction, IEEE Press, Tokyo, Japan. p. 131–132.

Hegel, F., 2012. Effects of a robot's aesthetic design on the attribution of social capabilities, in: 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, IEEE. doi:10.1109/roman.2012.6343796.

Hägele, M., Nilsson, K., Pires, J.N., 2008. Industrial Robotics, in: Siciliano, B., Khatib, O. (Eds.), Springer Handbook of Robotics. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 963–986. doi:10.1007/978-3-540-30301-5_43.

Kim, B., Bruce, M., Brown, L., de Visser, E., Phillips, E., 2020. A comprehensive approach to validating the uncanny valley using the anthropomorphic RoBOT (ABOT) database, in: 2020 Systems and Information Engineering Design Symposium (SIEDS), IEEE. doi:10.1109/sieds49339.2020.9106675.

Koenig, A.M., 2018. Comparing prescriptive and descriptive gender stereotypes about children, adults, and the elderly. Frontiers in Psychology 9. doi:10.3389/fpsyg.2018.01086.

Kopp, T., Baumgartner, M., Kinkel, S., 2022. "it's not paul, it's a robot": The impact of linguistic framing and the evolution of trust and distrust in a collaborative robot during a human-robot interaction. SSRN Electronic Journal doi:10.2139/ssrn.4113811.

Kriz, S., Ferro, T.D., Damera, P., Porter, J.R., 2010. Fictional robots as a data source in HRI research: Exploring the link between science fiction and interactional expectations, in: 19th International Symposium in Robot and Human Interactive Communication, IEEE. doi:10.1109/roman.2010.5598620.

Ladwig, R.C., Ferstl, E.C., 2018. What's in a name?, in: Proceedings of the 4th Conference on Gender & IT, ACM Press. doi:10.1145/3196839.3196851.

Lampousaki, S., 2010. Stereotypes about gender and work .

Law, T., Chita-Tegmark, M., Scheutz, M., 2020. The interplay between emotional intelligence, trust, and gender in human–robot interaction. International Journal of Social Robotics 13, 297–309. doi:10.1007/s12369-020-00624-1.

Li, D., Rau, P.L.P., Li, Y., 2010. A cross-cultural study: Effect of robot appearance and task. International Journal of Social Robotics 2, 175–186. doi:10.1007/s12369-010-0056-9.

Van der Loos, H.M., Reinkensmeyer, D.J., 2008. Rehabilitation and Health Care Robotics, in: Siciliano, B., Khatib, O. (Eds.), Springer Handbook of Robotics. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1223–1251. doi:10.1007/978-3-540-30301-5_54.

Matheson, E., Minto, R., Zampieri, E.G.G., Faccio, M., Rosati, G., 2019. Human–robot collaboration in manufacturing applications: A review. Robotics 8, 100. doi:10.3390/robotics8040100.

Millo, F., Gesualdo, M., Fraboni, F., Giusino, D., 2021. Human likeness in robots: Differences between industrial and non-industrial robots, in: European Conference on Cognitive Ergonomics 2021, ACM. doi:10.1145/3452853.3452886.

Moulton, J., Robinson, G.M., Elias, C., 1978. Sex bias in language use: "neutral" pronouns that aren't. American Psychologist 33, 1032–1036. doi:10.1037/0003-066x.33.11.1032.

Onnasch, L., Hildebrandt, C.L., 2021. Impact of anthropomorphic robot design on trust and attention in industrial human-robot interaction. J. Hum.-Robot Interact. 11. doi:10.1145/3472224.

Onnasch, L., Roesler, E., 2019. Anthropomorphizing robots: The effect of framing in human-robot collaboration, SAGE Publications Inc. pp. 1311–1315. doi:10.1177/1071181319631209.

Onnasch, L., Roesler, E., 2021. A taxonomy to structure and analyze human-robot interaction. International Journal of Social Robotics 13, 833–849. doi:10.1007/s12369-020-00666-5.

Paetzel, M., Peters, C., Nyström, I., Castellano, G., 2016. Congruency matters - how ambiguous gender cues increase a robot's uncanniness, in: Social Robotics. Springer International Publishing, pp. 402–412. doi:10.1007/978-3-319-47437-3_39.

Perugia, G., Guidi, S., Bicchi, M., Parlangeli, O., 2022. The shape of our bias: Perceived age and gender in the humanoid robots of the abot database, in: Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction, IEEE Press, Sapporo, Hokkaido, Japan. p. 110–119.

Phillips, E., Zhao, X., Ullman, D., Malle, B.F., 2018. What is human-like?: Decomposing robots' human-like appearance using the anthropomorphic robot (abot) database, in: 2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 105–113.

Powers, A., Kramer, A., Lim, S., Kuo, J., lai Lee, S., Kiesler, S., 2005. Eliciting information from people with a gendered humanoid robot, in: ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, IEEE. doi:10.1109/roman.2005.1513773.

Reich-Stiebert, N., Eyssel, F., 2017. (ir)relevance of gender?, in: Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, ACM. doi:10.1145/2909824.3020242.

Reilly, K., 2019. Robots and anthropomorphism in science-fiction theatre: From rebellion to domesticity and back again, in: Wynants, N. (Ed.), Media Archaeology and Intermedial Performance: Deep Time of the Theatre. Springer International Publishing, Cham, pp. 193–210. doi:10.1007/978-3-319-99576-2_9.

Robotics, R., n.d. URL: https://www.rethinkrobotics.com/sawyer. retrieved July 18, 2022, from Sawyer collaborative robots for Industrial Automation. Sawyer, the high performance collaborative robot | Rethink Robotics. (n.d.). Retrieved July 18, 2022, from https://www.rethinkrobotics.com/sawyer.

Roesler, E., Manzey, D., Onnasch, L., 2021. A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction. Science Robotics 6, eabj5425. doi:10.1126/scirobotics.abj5425.

Roesler, E., Naendrup-Poell, L., Manzey, D., Onnasch, L., 2022. Why context matters: The influence of application domain on preferred degree of anthropomorphism and gender attribution in human-robot interaction. International Journal of Social Robotics doi:10.1007/s12369-021-00860-z.

Roesler, E., Onnasch, L., Majer, J.I., 2020. The effect of anthropomorphism and failure comprehensibility on human-robot trust, SAGE Publications Inc. pp. 107–111. doi:10.1177/1071181320641028.

Stadler, S., Weiss, A., Mirnig, N., Tscheligi, M., 2013. Anthropomorphism in the Factory: A Paradigm Change?, in: Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction, IEEE Press. pp. 231–232. doi:10.1109/HRI.2013.6483586.

Sung, J.Y., Guo, L., Grinter, R.E., Christensen, H.I., 2007. "my roomba is rambo": Intimate home appliances, in: UbiComp 2007: Ubiquitous Computing. Springer Berlin Heidelberg, pp. 145–162. doi:10.1007/978-3-540-74853-3_9.

Tay, B., Jung, Y., Park, T., 2014. When stereotypes meet robots: The double-edge sword of robot gender and personality in human–robot interaction. Computers in Human Behavior 38, 75–84. doi:10.1016/j.chb.2014.05.014.

Waytz, A., Cacioppo, J., Epley, N., 2010. Who sees human? the stability and importance of individual differences in anthropomorphism. Perspectives on psychological science : a journal of the Association for Psychological Science 5, 219–232. doi:10.1177/1745691610369336.

Złotowski, J., Khalil, A., Abdallah, S., 2020. One robot doesn't fit all: aligning social robot appearance and job suitability from a middle eastern perspective. AI &SOCIETY 35, 485–500. doi:10.1007/s00146-019-00895-x.

3 General Discussion

The purpose of this thesis was to gain a better understanding of the effectiveness and desirability of anthropomorphism in HRI. The results of the first study provide supporting evidence that anthropomorphism is no universal remedy to facilitate HRI. Moreover, as the utility of anthropomorphism could be shown to be highly dependent on the field of application in which robots were implemented. This thesis aimed to further investigate whether also people's preferences for robot design depend on the application domain and sociability of the task. First, therefore, the form-function and function-form relationship of anthropomorphic design will be discussed in 3.1.

In addition, as robot design might reinforce existing gender stereotypes of different application domains, gender associations were examined. The studies of this thesis showed the challenges of designing robots gender-neutrally, as well as a transfer of stereotypes to even gender-neutral robots, especially in the industrial domain. This will be discussed in detail in section 3.2.

Of course, the results of the studies incorporated in this thesis should be interpreted by having the limitations in mind. The methodological limitations will be discussed in section 3.3. Lastly, promising starting points for future research will be presented in section 3.4.

3.1 The double-edged Sword of Anthropomorphism

In general, the meta-analysis served as a systematic approach to incorporate conflicting results of earlier research which showed positive (e.g., Moon et al., 2014), as well as negative effects of anthropomorphism by design (e.g., Onnasch & Roesler, 2019) apart from the uncanny valley (Mori et al., 2012). The key finding of the presented meta-analysis is that implementing anthropomorphic features leads overall to a positive effect. However, the effectiveness seems to depend considerably on factors of robot morphology and the interaction context described in the taxonomy of Onnasch and Roesler (2021). On part of the robot, it can be stated that multiple and communicational features of anthropomorphism are the most effective ones. The effect that multiple aspects were especially effective is consistent with the claim that anthropomorphism evokes expectations in regard to functionality (Lohse et al., 2007). Having multiple anthropomorphic design features might increase the likelihood that expectations are met, for example via a match of anthropomorphic appearance and anthropomorphic communication (Haring et al., 2018; Klüber & Onnasch, 2022). In addition, communicational implementations often fulfill a functional purpose, which seems to make them effective (Bonarini, 2020). Moreover, only task-relevant implementations like anthropomorphic movements (e.g., Kuz et al., 2013; Mayer et al., 2013) led to an improvement on a behavioral level.

Besides the robot itself, also other factors of the broader interaction frame were found to be important. In particular, the methodology used in different studies seems to be decisive for the outcomes of anthropomorphism. The results support the hypothesis that the embodiment of robots matters for perception and behavior in HRI (Deng et al., 2019; Wainer et al., 2006; Wainer et al., 2007). Whereas positive effects of anthropomorphism were revealed for subjective and objective outcomes for research using embodied robots, only effects on subjective measures emerged for depicted robots (Roesler et al., 2022). This will be further discussed in regard to the studies presented in this thesis in section 3.3. Apart from those aspects, there is one key finding of the meta-analysis, which motivated the subsequent online studies, namely that the effectiveness of anthropomorphism is domain specific.

Specifically, anthropomorphism mostly facilitated aspects of social HRI. On a detailed level, the analysis of outcome variables showed a lack of evidence for task-related variables like perceived safety and task performance, which are typically assessed in other domains than the social one (Steinfeld et al., 2006). Second, more generally, only HRI taking place in the social application domain was consistently showing positive consequences of anthropomorphism. In contrast, the service domain did not benefit at all, and the industrial domain just partially was affected from the implementation of anthropomorphic features. The different effectiveness of anthropomorphism in different application domains might be associated with the different challenges that occur in the respective interaction (Sheridan, 2016). In contrast to the industrial and service domain, social interaction and not productivity is the main goal in the social domain (Bartneck & Forlizzi, 2004). Whereas in the social domain, the emphasis is on communication and emotion (Breazeal, 2003; Duffy, 2003), in the industrial and service domain, the physical dependability and safety are the major concerns (De Santis et al., 2008). In my view, the expectations in regard to social stimulation of the respective application domains are the most compelling explanation for the present set of findings. For this reason, purely decorative anthropomorphism might be beneficial in the social domain, as those seem to meet the expectations of participants. In contrast, decorative anthropomorphism seems to be ineffective or possibly negative in the more task-related domains (Onnasch & Hildebrandt, 2021; Onnasch & Roesler, 2019; Roesler et al., 2021).

To investigate the assumption that anthropomorphic design which is not implemented in a functional but purely decorative way, is desired mainly in the social domain, the second study of this thesis was conducted. It is striking that the findings of the second study broadly support that assumption. If participants had a choice between different anthropomorphic robots that did not differ in their functionality, they chose most often robots with a high level of anthropomorphism in the social domain. Vice versa, participants preferred robots with low levels of anthropomorphism in the industrial domain. Interestingly, no clear preference was present for the service domain. Taken together, these results indicate that in line with the *matching hypotheses* the appearance of the robot should match the task (Goetz et al., 2003; Złotowski et al., 2020), and in this case the overall application domain. Compared to earlier research (Goetz et al., 2003; Li et al., 2010; Złotowski et al., 2020), this study is as best to my knowledge, the first which shed light on the domain rather than specific tasks. This might also explain why there were such wide-ranging preferences from low to high anthropomorphic robots in the service domain. The service domain includes a variety of tasks (Złotowski et al., 2020) which could have been represented in the participant's mind when choosing a robot. Our description contained non-social services like transport and cleaning, as well as social services like having contact with customers. Depending on what participants put more weight on, they might have chosen the associated degree of anthropomorphism in regard to the required sociability of the task. Also, the findings of the social and industrial domains might be explained by the associated sociability of the task, as both were described and are generally associated with opposing degrees of sociability.

This assumption directly led to the research question addressed in the fourth study. The objective of this study was to investigate what is more crucial for the preferred degree of anthropomorphism, the application domain, or the sociability of the task. The pattern of results of this study suggests that task sociability is an important factor in the social domain. This is consistent with research focusing on the *matching hypothesis* in the social domain (Goetz et al., 2003; Złotowski et al., 2020). However, in the industrial domain independent of the sociability of the task, robots with a low degree of anthropomorphism were preferred. This result stands in contrast to the current development of the market, where a general trend for anthropomorphic shapes can be found, even though there is sometimes no functional need for them (Baratta, 2015). Millo et al. (2021) showed that industrial robots currently fully or partially used in industry incorporate significantly fewer anthropomorphic appearance features than robots that are used non-work related. However, the mean values of human likeness (Phillips et al., 2018) of fully (M=31.9) and partially (M=33.4) industrial used robots are still considerably higher than the low anthropomorphic robots used in the second (M=9.14) and fourth study (M=7.28) of this thesis. Currently, industrial robots seem to correspond rather to the medium and high anthropomorphic robots used in this thesis in regard to human likeness. However, robots of these categories were not preferred in the industrial domain.

The results of the current work thus challenge the trend of equipping industrial robots with decorative anthropomorphic features. This is an important finding in the understanding of anthropomorphism in domains apart from the social one.

Taken together, both the meta-analysis and the two discussed online studies show that drawing conclusions from tasks that require high sociability is not suitable for all collaborative HRI scenarios. Close collaboration does not automatically benefit from more anthropomorphic appearance. In conclusion, decorative anthropomorphic features should only be adopted in tasks that aim to cognitively, emotionally, or physically stimulate social interaction. But even in these interactions, anthropomorphism can have negative consequences in regard to social categorization, as discussed in the next section 3.2.

3.2 Assigning Gender to (Gender-Neutral) Robots

The second superordinate objective of this thesis was to investigate the possible transfer of occupational gender stereotypes to robots. In general, the results of the implicit association test of the second study indicate that stereotypical gender-occupation associations already exist in HRI. That is, the association of female-designed robots and the social domain, as well as male-designed robots and the industrial domain, was stronger than diametrical pairings. This result already illustrates the risk that gendering robots by their design according to human stereotypes might reinforce those stereotypes. This shows that gender-neutral robots might offer the opportunity to avoid ethically questionable design. In contrast to past research that had mainly focused on the consequences of intentionally gendered robots (e.g., Chita-Tegmark et al., 2019; Eyssel & Hegel, 2012; Powers et al., 2005), the present study investigated robots without obvious gender cues. The good news is that gender-neutral robot design seems to be effective to reduce stereotypical associations. In general, all online studies of this thesis demonstrated a mainly gender-neutral or functional association of robots in all domains.

However, if robots were gendered, the results of the second study of this thesis were consistent with the claim of a general male-robot bias. This result could indicate a generally stronger association of technologies like robots with masculinity than with femininity (Dufour & Nihan, 2016; Faulkner, 2001). This might be related to more men currently closely working in the technological field (Lohan & Faulkner, 2004), or masculine trait stereotypes like physical strength and/or computational capabilities (Berg & Lie, 1995). It, therefore, seemed reasonable that also a general male-robot bias might exist. However, this conclusion would only be valid if the possible methodological drawbacks were eliminated. The two central methodological drawbacks included the influence of grammatical gender of the word robot in German and the possible influence of appearance, even though no obvious gender cues were present. The results of this study provided no evidence that the mother language of participants influenced the perceived gender. Nevertheless, the third study revealed that the grammatical gender (Boroditsky et al., 2003; Samuel et al., 2019) used in the robot introduction seemed to matter for the assigned gender. The neuter introduction used in the second study of this thesis ("the system") seemed to be a suitable approach to at least partially avoid a male-robot bias induced via language aspects. Moreover, consistent with a current study of Perugia et al. (2022) the results revealed that the robots of the high anthropomorphic group were all perceived as male. Having in mind that, especially in the social domain, robots with high anthropomorphic (masculine) robot design were chosen, the validity of the results of the second study can be questioned.

For this reason, the fourth study investigated the possible transfer of gender stereotypes to robots, which were clearly perceived as gender-neutral in regard to their appearance with an English sample. In contrast to the second study, the results did not confirm a general male-robot bias, possibly existing due to the association of technology and masculinity (Faulkner, 2001). Moreover, no one-on-one transfer of occupational stereotypes was revealed. There are two opposing key findings of the fourth study in regard to assigned gender. First, in the social domain, no binary gender association was significantly more often assigned than the other one. Second, in the industrial domain, especially for tasks requiring sociability, a male bias occurred. This result implies that even gender-neutral robots with low degrees of anthropomorphism can be categorized in regard to the occupational gender stereotype. Interestingly, deanthropomorphization might be used here as a countermeasure to this automatic categorization by framing the robot as a functional tool (e.g., with a functional name) (Kopp et al., 2022; Onnasch & Roesler, 2019) with grammatically neuter descriptions. A deliberate violation of the stereotype would also be conceivable by mismatching the stereotypical gender (Reich-Stiebert & Eyssel, 2017) with a female-gendered industrial robot. However, both approaches need to be considered in future research to evaluate their effectiveness.

Altogether, the findings of the studies investigating occupational gender stereotypes illustrated in line with other research (Perugia et al., 2022; Perugia & Lisy, 2022) how challenging gender-neutral or genderless robot design is, as soon as higher degrees of anthropomorphic appearance are implemented. Moreover, even gender-neutral robots do not completely prevent the transfer of occupational gender stereotypes.

3.3 Critical Considerations

Although the presented studies systematically investigated anthropomorphism in HRI, and especially the role of different interaction contexts, it is important to recognize several potential limitations. First, let me consider the limitations of the meta-analysis which represented the first and most general part of my research.

The greatest strength of a meta-analysis is also its greatest weakness: combining information across multiple studies. The metaphor "combining apples with oranges" is often used to illustrate this pitfall. Every meta-analysis depends on the quality of the studies which are included for data analysis (Greco et al., 2013), and even though we have tried to ensure a certain quality via the pre-registration, inclusion criteria, and measures of publication bias, it can not be ruled out completely that the current state-of-the-art research is biased. Moreover, in the meta-analysis statistical heterogeneity was clearly evident, and the moderators could not come close to fully explaining this heterogeneity. Probably the best way to address these points of criticism is to make meta-analytic data publically available. This allows different researchers to continuously and easily extend and reanalyze the data set (Lakens et al., 2016). A great strength of the presented meta-analysis is, therefore, that it can be updated. In addition, the synthesis of data did not only reveal quantitative data in terms of effect sizes but also concerning the current focus of attention, and consequently showed substantial research gaps, which will be discussed in section 3.4.

The biggest content-related problem that has already became apparent in the meta-analysis and has also played a major role in all other studies of this thesis is the quantifiability of anthropomorphism. Most commonly, anthropomorphism is measured from completely machinelike (0% anthropomorphic) to completely humanlike (100% anthropomorphic) (Bartneck et al., 2009; Phillips et al., 2018). So far, however, this has almost only referred to the appearance (Bartneck et al., 2009; Phillips et al., 2018) and very rarely to other aspects like movements, communication, or context. Even for appearance, the investigation with similar scale anchors comes up with different results. For example, the second and third studies of this thesis showed considerable differences compared to the score of the ABOT database (Phillips et al., 2018), even though both were measured on a scale from 0-100%. The values of the ABOT database (low M=9.14; medium M= 23.06; high M=49.20) were remarkably smaller compared to the values of the second study (low M=18.07; medium M= 44.17; high M=79.06). A comparable trend was observable by comparing the mean values of the ABOT database (low M=8.95; medium M= 23.39; high M=49.24) with the mean values of the third study (low M=14.20; medium M= 35.90; high M=68.80). This might be explained by the range of stimuli presented to the participants. In Phillips et al. (2018) study, each participant rated 50 robots, which could range from extremely technical robots like sphero (basically a robotic ball) (Faria et al., 2016) to extremely humanlike robots like geminoid h1-4 (a robotic replica of Hiroshi Ishiguro) (Nishio et al., 2007). The context in which each robot was presented to the participants seems to have influenced the perceived anthropomorphism. For example, the interaction of a low anthropomorphic robot with a higher anthropomorphic one can already change the perceived anthropomorphism of the robots (Ueno et al., 2019). Carry-over effects and contrast effects seem to influence the anthropomorphism of robots. It is therefore a big challenge to make a valid assumption and to objectively measure anthropomorphism. Moreover, there is a substantial need to validly measure anthropomorphism apart from the appearance in regard to movement, communication, and context. Certainly, no approach could have been used to assign specific values to each type of morphology to enable a graduated quantification of anthropomorphism. The meta-analysis in particular showed the immense need for such a universal metric.

Other limitations concern the set of three online studies which were conducted to follow up on some more specific issues of anthropomorphism in HRI. In these studies, only depicted robots could be used and already the results of the meta-analysis suggest that this might compromise the generalizability of the results to HRI with embodied robots. As stated in a follow-up analysis of the meta-analysis (Roesler et al., 2022) using depicted robots to investigate anthropomorphism showed positive effects on a subjective level but failed to show any effects on an objective level. Even more critical for the presented studies, anthropomorphism might lead to a more positive perception if robots are only depicted. This is related to the fact that expectations set via the anthropomorphic appearance can not be violated by for example loud and jerky movements (Trovato et al., 2018). However, this also illustrates how strong the effect of preferring low anthropomorphic robots in the industrial domain might be, as even potentially idealized anthropomorphic robots were chosen significantly less in this domain. But it also shows how important research with embodied robots is to draw conclusions about the actual interaction. Since the online studies were primarily focused on the general preferences of the public and conducted in times of the COVID-19 pandemia (Feil-Seifer et al., 2020), the online approach still seems justifiable.

Another limitation of the online studies presented in this thesis is the assessment of ascribed gender in general. Stereotypes are a socially sensitive issue (Grimm, 2010), and asking participants about their personal judgment (Koenig, 2018) of the robot's gender might have led to more socially desired answers. This effect seems to be particularly critical for the interpretation of the third and fourth studies. In both studies, the robots' perceived gender was overtly assessed by either a scale or a pronoun technique. As no forced choice between different binary genders was presented, the awareness that assigned gender was investigated might have led to the more gender-neutral and functional ascription. For this aspect, however, the online survey could have been advantageous, as social desirability is inversely related to the anonymity that a person experiences (Ben-Ze'ev, 2003; Joinson, 1999). Even though it might have been a subject of social desirability effects, the fact that the assessment of assigned gender was not binary is a great advantage of the current thesis. Especially, the introduction of a functional category allowed for broadening the idea of robot gender. The naming and pronoun technique have turned out as useful tools to investigate not only what gender was attributed to robots, but more generally, whether gender was attributed at all. Past researchers have used scales comparable to the scale used in the third study of this thesis, ranging from female to male to assess robot genderedness (e.g., Eyssel & Hegel, 2012; Ghazali et al., 2018; Kuchenbrandt et al., 2014). The extension of this methodological approach with the naming and pronoun technique allowed for a more subtle (Perugia & Lisy, 2022) and differentiated assessment of possible gender effects in HRI.

In sum, and despite the discussed limitations, the studies presented in this thesis provide novel insights and suggest several theoretical and practical implications. Most prominently, the results of the presented studies offer an invitation for future research, as discussed in the next section.

3.4 Research Outlook

The current thesis provides manifold perspectives for future research. Quite some topics can be derived from the results of the meta-analysis alone. In the course of this analysis, a major contribution may be that it raised a variety of intriguing questions for future studies. Moreover, it can be used to identify research gaps as well as overrepresented topics, which are currently setting the tone of HRI research. There are at least three key potential starting points revealed in this analysis.

A first starting point concerns the measures used in the current HRI studies to investigate the role of anthropomorphism. The assessment of perception was highly overrepresented compared to any other outcome (attitudes, affect, and behavior). Actually, the idea behind measuring perception in HRI was that especially the success of social robots is not only associated with performance measures which are typically decisive for industrial robots (Bartneck et al., 2009). Moreover, perception is a very effective measure concerning time, money, and implementation into the experimental setups (Bartneck et al., 2009; Weiss & Bartneck, 2015) in comparison to behavioral measures (Steinfeld et al., 2006). In addition, the popularity of perceptional measures is not only related to the aspect that questionnaires are easy to use. Attitudes like trust are also accessible via questionnaires (Chita-Tegmark et al., 2021) but investigated substantially less than perception. This might again be associated with the effort of enabling an actual interaction. Attitudes like trust are highly dynamic, and mostly from interest is the change in this attitude due to robot-related factors like anthropomorphism (Hancock et al., 2011). It is therefore nearly inevitable to have an interaction between robots and humans to investigate attitudes. In comparison, perceptions can be assessed static by evaluating for example a picture of a robot. The enormous pressure of researchers to publish quickly and in large quantities (Sarewitz, 2016) makes perceptional measures so attractive, as no other kind of measure is so easily implementable. However, this also shifts the focus from the most important aspect of HRI – the I. The actual interaction should always be the key objective of research, therefore the meta-analysis is a wake-up call for more interaction-focused HRI.

A second starting point for future research was revealed by the moderator analysis of the field of application. It is widely assumed that anthropomorphism facilitates collaboration (i.e., interdependence in temporal and spatial interaction) due to the shift from tool to team partner (Duffy, 2003; Phillips et al., 2011). However, the analysis revealed that this might not be the case apart from the social domain. Moreover, effects found in the social domain might not be transferrable to other domains. It is therefore important to investigate which effects of anthropomorphism (e.g., task-relevant implications) cross application domains, and which ones are exclusively effective for the social domain. Much research still needs to be done to form a common foundation of knowledge on how robots should look and act for different application domains and tasks (Dautenhahn, 2007).

Lastly, and as already discussed in section 3.3 there is a pronounced need in the HRI community to develop new measurements of anthropomorphism even going beyond appearance. For appearance, the assumption that the effectiveness of anthropomorphism depends on the degree of anthropomorphism is widely supported (Bartneck et al., 2007; Kim et al., 2020; Mori et al., 2012). In order to extend this assumption to other types of morphologies, we first of all, need a sufficient measurement tool for anthropomorphism.

I hope that the presented meta-analysis will stimulate further investigation of the effectiveness of anthropomorphism to move the often general assumption of benefit to a more fine-grained picture.

The important role of the field of application revealed in the meta-analysis was investigated in the second and fourth studies of this thesis. The studies changed the perspective from effectiveness to desirability to broaden the view to more general societal issues. In terms of future research, the fourth study already investigated the biggest open question of the second study – the interplay of the application domain and the sociability of the task. By integrating the results of both studies, it becomes apparent that the desirability of anthropomorphism depends on both aspects. As these studies focused completely on task-irrelevant anthropomorphic appearance, much work remains to be done in regard to other types of morphologies (i.e., movements, communication, and context) before a broader understanding of desirability can be established.

In addition to this, I suggest several avenues of research in regard to gender associations in HRI. The main novelty value of the three online studies lies in the fact that robots without obvious gender cues were investigated. In line with the results of Perugia et al. (2022) the second and third studies of this thesis showed that even robots without obvious gender cues can be perceived as rather male concerning their appearance. Moreover, this was particularly the case with increasing degrees of anthropomorphism. Since anthropomorphism is so strongly connected to gendering robots, the general question of whether this is desirable needs to be approached. It seems promising to use the tabula rasa of clearly gender-neutral robots in regard to appearance. However, this tabula rasa can be filled with a gendered perception due to grammatical gender and occupational stereotypes, as the third and fourth studies illustrated. There is still a need for research that explores what aspects can accidentally fill this tabula rasa. For future robot designs, the challenge remains not only whether we could come close to gender-neutral anthropomorphic robots (Perugia & Lisy, 2022), but also how we could maintain this gender-neutrality throughout the interaction. A particular challenge will be the implementation of gender-neutral communication which matches the appearance of the robot (Yu et al., 2022). In terms of future research, it would be useful to extend the current findings by investigating the effects of gender-neutral robots in regard to different morphologies in real interactions.

In summary, research should build the fundament for the design decisions on whether anthropomorphism should be implemented. Whereas this thesis focused on the danger of replicating occupational gender stereotypes, anthropomorphism also entails other risks like the transfer of racial stereotypes, the development of unsuitable emotional attachment, or the possibility for emotional manipulation (Darling, 2015). The robot design should therefore not take anthropomorphism as the main goal, but rather as a carefully and exceptionally implementable feature, whenever it fundamentally contributes to a pleasant and smooth interaction. Possibly, it could also be useful to completely refrain from purely decorative anthropomorphic features to avoid ethical issues whenever possible. Future research needs to investigate what alternatives can be used besides anthropomorphism to make HRI more intuitive.

3.5 Conclusion

The objective of this thesis was to challenge anthropomorphism as a design element in regard to effectiveness, desirability, and possible transfer of occupational gender stereotypes. Based on a meta-analysis and three consecutive online studies, it can be concluded that the effectiveness and desirability of anthropomorphism highly depend on the application domain in which the HRI takes place. The anthropomorphic design was most effective in the social domain, whereas the results for the service domain and industrial domain were at least mixed. This was also reflected in the desired degree of anthropomorphism, which was higher for the social domain, lower for the industrial domain, and mixed for the service domain. Moreover, the sociability of the task was decisive for the desired degree of anthropomorphism in the social but not industrial domain. In addition, gender-neutral robots primarily elicited a neutral or functional association. However, if robots were gendered, a replication of occupational stereotypes with more male gender associations was present in the industrial domain. No differences in binary gender associations were revealed in the social domain.

Despite the discussed limitations, all studies presented in this thesis argue against the conventional wisdom that anthropomorphism in general facilitates the interaction and is desirable in HRI. Rather, the form should be in accordance with the functionality, and decorative anthropomorphism should be avoided whenever possible. So the answer to the question if anthropomorphic design is beneficial and desirable in HRI, the answer is the same as to the most questions posed in psychological research – it depends.

References

Adachi, T. (2013). Occupational gender stereotypes: Is the ratio of women to men a powerful determinant? *Psychological Reports*, *112*(2), 640–650. https://doi.org/10.2466/17.07.pr0.112.2.640-650

Alesich, S., & Rigby, M. (2017). Gendered robots: Implications for our humanoid future. *IEEE Technology and Society Magazine*, *36*(2), 50–59. https://doi.org/10.1109/mts.2017.2696598

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, *41*(3), 258–290. https://doi.org/10.1037/h0055756

Baratta, D. (2015). Industrial collaborative robot design. *Paper presented at the CEUR Workshop Proceedings*, 2–8.

Bartneck, C., & Forlizzi, J. (2004). A design-centred framework for social human-robot interaction. *Proceedings of the 2004 IEEE International Workshop on Robot and Human Interactive Communication*. https://doi.org/10.1109/roman.2004.1374827

Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2007). Is the uncanny valley an uncanny cliff? *Proceedings of the 2007 IEEE International Symposium on Robot and Human Interactive Communication*. https://doi.org/10.1109/roman.2007.4415111

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Ben-Ze'ev, A. (2003). Privacy, emotional closeness, and openness in cyberspace. *Computers in Human Behavior*, *19*(4), 451–467. https://doi.org/10.1016/s0747-5632(02)00078-x

Berg, A.-J., & Lie, M. (1995). Feminism and constructivism: Do artifacts have gender? *Science, Technology, & Human Values*, *20*(3), 332–351.

Bernotat, J., Eyssel, F., & Sachse, J. (2017). Shape it – the influence of robot body shape on gender perception in robots. In *Social robotics* (pp. 75–84). Springer International Publishing. https://doi.org/10.1007/978-3-319-70022-9_8

Bernotat, J., Eyssel, F., & Sachse, J. (2021). The (fe)male robot: How robot body shape impacts first impressions and trust towards robots. *International Journal of Social Robotics*, *13*(3), 477–489. https://doi.org/10.1007/s12369-019-00562-7

Bonarini, A. (2020). Communication in human-robot interaction. *Current Robotics Reports*, *1*(4), 279–285. https://doi.org/10.1007/s43154-020-00026-1

Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. *Language in mind: Advances in the study of language and thought*, *22*, 61–79.

Breazeal, C. (2003). Toward sociable robots. *Robotics and Autonomous Systems*, *42*(3), 167–175. https://doi.org/10.1016/S0921-8890(02)00373-1

Bryant, D., Borenstein, J., & Howard, A. (2020). Why should we gender? *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. https://doi.org/10.1145/3319502.3374778

Carothers, B. J., & Reis, H. T. (2013). Men and women are from earth: Examining the latent structure of gender. *Journal of Personality and Social Psychology*, *104*(2), 385–407. https://doi.org/10.1037/a0030437

Chaminade, T., & Okka, M. M. (2013). Comparing the effect of humanoid and human face for the spatial orientation of attention. *Frontiers in Neurorobotics*, *7*. https://doi.org/10.3389/fnbot.2013.00012

Chita-Tegmark, M., Law, T., Rabb, N., & Scheutz, M. (2021). Can you trust your trust measure? *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. https://doi.org/10.1145/3434073.3444677

Chita-Tegmark, M., Lohani, M., & Scheutz, M. (2019). Gender effects in perceptions of robots and humans with varying emotional intelligence. *Proceedings of* *the 2019 ACM/IEEE International Conference on Human-Robot Interaction.* https://doi.org/10.1109/hri.2019.8673222

Darling, K. (2015). 'who's johnny?' anthropomorphic framing in human-robot interaction, integration, and policy. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.2588669

Darling, K., Nandy, P., & Breazeal, C. (2015). Empathic concern and the effect of stories in human-robot interaction. *Proceedings of the 2015 IEEE International Symposium on Robot and Human Interactive Communication.* https://doi.org/10.1109/roman.2015.7333675

Dautenhahn, K. (2007). Methodology &amp themes of human-robot interaction: A growing research field. *International Journal of Advanced Robotic Systems*, *4*(1), 15. https://doi.org/10.5772/5702

De Santis, A., Siciliano, B., De Luca, A., & Bicchi, A. (2008). An atlas of physical human–robot interaction. *Mechanism and Machine Theory*, *43*(3), 253–270. https://doi.org/https://doi.org/10.1016/j.mechmachtheory.2007.03.003

Deng, E., Mutlu, B., & Mataric, M. (2019). Embodiment in socially interactive robots. *arXiv preprint arXiv:1912.00312.*

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3

Dufour, F., & Nihan, C. E. (2016). Do robots need to be stereotyped? technical characteristics as a moderator of gender stereotyping. *Social Sciences*, *5*(3). https://doi.org/10.3390/socsci5030027

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*(1), 109–128. https://doi.org/10.1037/0033-2909.110.1.109

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

Faria, M., Costigliola, A., Alves-Oliveira, P., & Paiva, A. (2016). Follow me: Communicating intentions with a spherical robot. *Proceedings of the 2016 IEEE International Symposium on Robot and Human Interactive Communication*. https://doi.org/10.1109/roman.2016.7745189

Faulkner, W. (2001). The technology question in feminism. *Women's Studies International Forum*, *24*(1), 79–95. https://doi.org/10.1016/s0277-5395(00)00166-7

Feil-Seifer, D., Haring, K. S., Rossi, S., Wagner, A. R., & Williams, T. (2020). Where to next? the impact of covid-19 on human-robot interaction research. *ACM Transactions on Human-Robot Interaction*, *10*(1). https://doi.org/10.1145/3405450

Ferland, F., Létourneau, D., Aumont, A., Frémy, J., Legault, M.-A., Lauria, M., & Michaud, F. (2013). Natural interaction design of a humanoid robot. *Journal of Human-Robot Interaction*, *1*(2). https://doi.org/10.5898/jhri.1.2.ferland

Fink, J. (2012). Anthropomorphism and human likeness in the design of robots and human-robot interaction. In S. S. Ge, O. Khatib, J.-J. Cabibihan, R. Simmons, & M.-A. Williams (Eds.), *Social robotics* (pp. 199–208). Springer Berlin Heidelberg.

Fischer, K. (2022). Tracking anthropomorphizing behavior in human-robot interaction. *J. Hum.-Robot Interact.*, *11*(1). https://doi.org/10.1145/3442677

Fitzpatrick, P., Harada, K., Kemp, C. C., Matsumoto, Y., Yokoi, K., & Yoshida, E. (2016). Humanoids. In *Springer handbook of robotics* (pp. 1789–1818). Springer International Publishing. https://doi.org/10.1007/978-3-319-32552-1_67

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, *42*(3-4), 143–166. https://doi.org/10.1016/s0921-8890(02)00372-x

Galton, F. (1878). Composite portraiture. In *Inquiries into human faculty and its development.* (pp. 8–19). MacMillan Co. https://doi.org/10.1037/14178-003

Ghazali, A. S., Ham, J., Barakova, E. I., & Markopoulos, P. (2018). Effects of robot facial characteristics and gender in persuasive human-robot interaction. *Frontiers in Robotics and AI*, *5*. https://doi.org/10.3389/frobt.2018.00073

Goetz, J., Kiesler, S., & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation. *Proceedings of the 2003 IEEE International Workshop on Robot and Human Interactive Communication*, 55–60.

Greco, T., Zangrillo, A., Biondi-Zoccai, G., & Landoni, G. (2013). Meta-analysis: Pitfalls and hints. *Heart, lung and vessels*, *5*, 219–225.

Grimm, P. (2010). Social desirability bias. *Wiley international encyclopedia of marketing.*

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(5), 517–527. https://doi.org/10.1177/0018720811417254

Haring, K. S., Silvera-Tawil, D., Takahashi, T., Watanabe, K., & Velonaki, M. (2016). How people perceive different robot types: A direct comparison of an android, humanoid, and non-biomimetic robot. *Proceedings of the 2016 International Conference on Knowledge and Smart Technology.* https://doi.org/10.1109/kst.2016.7440504

Haring, K. S., Watanabe, K., Velonaki, M., Tossell, C. C., & Finomore, V. (2018). FFAB—the form function attribution bias in human–robot interaction. *IEEE*

Transactions on Cognitive and Developmental Systems, *10*(4), 843–851. https://doi.org/10.1109/tcds.2018.2851569

Haring, K. S., Watanabe, K., & Mougenot, C. (2013). The influence of robot appearance on assessment. *Proceedings of the 2013 ACM/IEEE International Conference on Human-Robot Interaction*, 131–132.

Hegel, F. (2012). Effects of a robot's aesthetic design on the attribution of social capabilities. *Proceedings of the 2012 IEEE International Symposium on Robot and Human Interactive Communication*. https://doi.org/10.1109/roman.2012.6343796

Hehman, E., Stolier, R. M., Freeman, J. B., Flake, J. K., & Xie, S. Y. (2019). Toward a comprehensive model of face impressions: What we know, what we do not, and paths forward. *Social and Personality Psychology Compass*, *13*(2), e12431. https://doi.org/10.1111/spc3.12431

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, *57*(4), 657–674. https://doi.org/10.1111/0022-4537.00234

Jackson, R. B., Williams, T., & Smith, N. (2020). Exploring the role of gender in perceptions of robotic noncompliance. *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. https://doi.org/10.1145/3319502.3374831

Joinson, A. (1999). Social desirability, anonymity, and internet-based questionnaires. *Behavior Research Methods, Instruments, &amp Computers*, *31*(3), 433–438. https://doi.org/10.3758/bf03200723

Keay, A., & Graduand, M. (2011). Emergent phenomena of robot competitions: Robot identity construction and naming. *Advanced Robotics and its Social Impacts*. https://doi.org/10.1109/arso.2011.6301972

Kim, B., Bruce, M., Brown, L., de Visser, E., & Phillips, E. (2020). A comprehensive approach to validating the uncanny valley using the anthropomorphic

RoBOT (ABOT) database. *2020 Systems and Information Engineering Design Symposium (SIEDS)*. https://doi.org/10.1109/sieds49339.2020.9106675

Klüber, K., & Onnasch, L. (2022). Appearance is not everything-preferred feature combinations for care robots. *Computers in Human Behavior*, *128*, 107128.

Koenig, A. M. (2018). Comparing prescriptive and descriptive gender stereotypes about children, adults, and the elderly. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01086

Kopp, T., Baumgartner, M., & Kinkel, S. (2022). "it's not paul, it's a robot": The impact of linguistic framing and the evolution of trust and distrust in a collaborative robot during a human-robot interaction. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4113811

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. *International Journal of Social Robotics*, *6*(3), 417–427. https://doi.org/10.1007/s12369-014-0244-0

Kuz, S., Mayer, M. P., Müller, S., & Schlick, C. M. (2013). Using anthropomorphism to improve the human-machine interaction in industrial environments (part i). In *Digital human modeling and applications in health, safety, ergonomics, and risk management. human body modeling and ergonomics* (pp. 76–85). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-39182-8_9

Ladwig, R. C., & Ferstl, E. C. (2018). What's in a name? *Proceedings of the 4th Conference on Gender & IT*. https://doi.org/10.1145/3196839.3196851

Lakens, D., Hilgard, J., & Staaks, J. (2016). On the reproducibility of meta-analyses: Six practical recommendations. *BMC Psychology*, *4*(1). https://doi.org/10.1186/s40359-016-0126-3

Lampousaki, S. (2010). Stereotypes about gender and work. *European Foundation for the Improvement of Living and Working Conditions*.

Lemaignan, S., Fink, J., Dillenbourg, P., & Braboszcz, C. (2014). The cognitive correlates of anthropomorphism. *2014 Human-Robot Interaction Conference, Workshop" HRI: a bridge between Robotics and Neuroscience"*, (CONF).

Li, D., Rau, P. L. P., & Li, Y. (2010). A cross-cultural study: Effect of robot appearance and task. *International Journal of Social Robotics*, *2*(2), 175–186. https://doi.org/10.1007/s12369-010-0056-9

Lohan, M., & Faulkner, W. (2004). Masculinities and technologies. *Men and Masculinities*, *6*(4), 319–329. https://doi.org/10.1177/1097184x03260956

Lohse, M., Hegel, F., Swadzba, A., Rohlfing, K., Wachsmuth, S., & Wrede, B. (2007). What can i do for you? appearance and application of robots. *Workshop on The Reign of Catz and Dogz? The role of virtual creatures in a computerised society*, 121–126.

Lugrin, B., Strole, E., Obremski, D., Schwab, F., & Lange, B. (2020). What if it speaks like it was from the village? effects of a robot speaking in regional language variations on users' evaluations. *Proceedings of the 2020 IEEE International Conference on Robot and Human Interactive Communication*. https://doi.org/10.1109/ro-man47096.2020.9223432

Mayer, M. P., Kuz, S., & Schlick, C. M. (2013). Using anthropomorphism to improve the human-machine interaction in industrial environments (part II). In *Digital human modeling and applications in health, safety, ergonomics, and risk management. human body modeling and ergonomics* (pp. 93–100). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-39182-8_11

Millo, F., Gesualdo, M., Fraboni, F., & Giusino, D. (2021). Human likeness in robots: Differences between industrial and non-industrial robots. *European Conference on Cognitive Ergonomics 2021*. https://doi.org/10.1145/3452853.3452886

Moon, A. J., Troniak, D. M., Gleeson, B., Pan, M. K., Zheng, M., Blumer, B. A., MacLean, K., & Crof, E. A. (2014). Meet me where i'm gazing: How shared attention gaze affects human-robot handover timing. *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, 334–341.

Mori, M., MacDorman, K., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics &amp Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/mra.2012.2192811

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nishio, S., Ishiguro, H., & Hagita, N. (2007). Geminoid: Teleoperated android of an existing person. *Humanoid robots: New developments*, *14*(343-352), 10–1109.

Nomura, T. (2017). Robots and gender. *Gender and the Genome*, *1*(1), 18–26. https://doi.org/10.1089/gg.2016.29002.nom

Onnasch, L., & Hildebrandt, C. L. (2021). Impact of anthropomorphic robot design on trust and attention in industrial human-robot interaction. *J. Hum.-Robot Interact.*, *11*(1). https://doi.org/10.1145/3472224

Onnasch, L., Kostadinova, E., & Schweidler, P. (2022). Humans can't resist robot eyes-reflexive cueing with pseudo-social stimuli. *Frontiers in Robotics and AI*, 72.

Onnasch, L., & Roesler, E. (2019). Anthropomorphizing robots: The effect of framing in human-robot collaboration. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 1311–1315. https://doi.org/10.1177/1071181319631209

Onnasch, L., & Roesler, E. (2021). A taxonomy to structure and analyze human-robot interaction. *International Journal of Social Robotics*, *13*(4), 833–849. https://doi.org/10.1007/s12369-020-00666-5

Persson, P., Laaksolahti, J., & Lönnqvist, P. (2000). Anthropomorphism-a multi-layered phenomenon. *Proceedings of Socially Intelligent Agents -The Human in the Loop, AAAI Press, Technical report FS-00-04*, 131–135.

Perugia, G., Guidi, S., Bicchi, M., & Parlangeli, O. (2022). The shape of our bias: Perceived age and gender in the humanoid robots of the abot database. *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, 110–119.

Perugia, G., & Lisy, D. (2022). Robot's gendering trouble: A scoping review of gendering humanoid robots and its effects on hri. https://doi.org/10.48550/ARXIV.2207.01130

Phillips, E., Ososky, S., Grove, J., & Jentsch, F. (2011). From tools to teammates: Toward the development of appropriate mental models for intelligent robots. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *55*(1), 1491–1495. https://doi.org/10.1177/1071181311551310

Phillips, E., Zhao, X., Ullman, D., & Malle, B. F. (2018). What is human-like?: Decomposing robots' human-like appearance using the anthropomorphic robot (abot) database. *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 105–113.

Powers, A., Kramer, A., Lim, S., Kuo, J., Lee, S.-l., & Kiesler, S. (2005). Eliciting information from people with a gendered humanoid robot. *Proceedings of the 2005 IEEE International Workshop on Robot and Human Interactive Communication*. https://doi.org/10.1109/roman.2005.1513773

Reich-Stiebert, N., & Eyssel, F. (2017). (ir)relevance of gender? *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. https://doi.org/10.1145/2909824.3020242

Riek, L. D., Rabinowitch, T.-C., Bremner, P., Pipe, A. G., Fraser, M., & Robinson, P. (2010). Cooperative gestures: Effective signaling for humanoid robots. *Proceedings of the 2010 ACM/IEEE International Conference on Human-Robot Interaction*. https://doi.org/10.1109/hri.2010.5453266

Robertson, J. (2010). Gendering humanoid robots: Robo-sexism in japan. *Body &amp Society*, *16*(2), 1–36. https://doi.org/10.1177/1357034x10364767

Roesler, E., Manzey, D., & Onnasch, L. (2021). A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction. *Science Robotics*, *6*(58), eabj5425. https://doi.org/10.1126/scirobotics.abj5425

Roesler, E., Manzey, D., & Onnasch, L. (2022). Embodiment matters in social hri research: Effectiveness of anthropomorphism on subjective and objective outcomes. *ACM Transactions on Human-Robot Interaction*.

Roesler, E., Onnasch, L., & Majer, J. I. (2020). The effect of anthropomorphism and failure comprehensibility on human-robot trust. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *64*(1), 107–111. https://doi.org/10.1177/1071181320641028

Samuel, S., Cole, G., & Eacott, M. J. (2019). Grammatical gender and linguistic relativity: A systematic review. *Psychonomic Bulletin &amp Review*, *26*(6), 1767–1786. https://doi.org/10.3758/s13423-019-01652-3

Sarewitz, D. (2016). The pressure to publish pushes down quality. *Nature*, *533*(7602), 147–147. https://doi.org/10.1038/533147a

Seaborn, K., & Pennefather, P. (2022a). Gender neutrality in robots: An open living review framework. *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, 634–638.

Seaborn, K., & Pennefather, P. (2022b). Neither "hear" nor "their": Interrogating gender neutrality in robots. *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, 1030–1034.

Sheridan, T. B. (2016). Human-robot interaction: Status and challenges. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *58*(4), 525–532. https://doi.org/10.1177/0018720816644364

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., & Goodrich, M. (2006). Common metrics for human-robot interaction. *Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction - HRI '06*. https://doi.org/10.1145/1121241.1121249

Tay, B., Jung, Y., & Park, T. (2014). When stereotypes meet robots: The double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior*, *38*, 75–84. https://doi.org/10.1016/j.chb.2014.05.014

Trovato, G., Paredes, R., Balvin, J., Cuellar, F., Thomsen, N. B., Bech, S., & Tan, Z.-H. (2018). The sound or silence: Investigating the influence of robot noise on proxemics. *Proceedings of the 2018 IEEE International Symposium on Robot and Human Interactive Communication*. https://doi.org/10.1109/roman.2018.8525795

Ueno, A., Hayashi, K., & Mizuuchi, I. (2019). Impression change on nonverbal non-humanoid robot by interaction with humanoid robot. *Proceedings of the 2019 IEEE International Conference on Robot and Human Interactive Communication*. https://doi.org/10.1109/ro-man46459.2019.8956240

Wainer, J., Feil-seifer, D., Shell, D., & Mataric, M. (2006). The role of physical embodiment in human-robot interaction. *Proceedings of the 2006 IEEE International Symposium on Robot and Human Interactive Communication*. https://doi.org/10.1109/roman.2006.314404

Wainer, J., Feil-Seifer, D. J., Shell, D. A., & Mataric, M. J. (2007). Embodiment and human-robot interaction: A task-based perspective. *Proceedings of the 2007 IEEE International Symposium on Robot and Human Interactive Communication*. https://doi.org/10.1109/roman.2007.4415207

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? the stability and importance of individual differences in anthropomorphism. *Perspectives on psychological science : a journal of the Association for Psychological Science*, *5*(24839457), 219–232. https://doi.org/10.1177/1745691610369336

Weiss, A., & Bartneck, C. (2015). Meta analysis of the usage of the godspeed questionnaire series. *Proceedings of the 2015 IEEE International Symposium on Robot and Human Interactive Communication*. https://doi.org/10.1109/roman.2015.7333568

White, M. J., & White, G. B. (2006). Implicit and explicit occupational gender stereotypes. *Sex Roles*, *55*(3), 259–266. https://doi.org/10.1007/s11199-006-9078-z

Wiese, E., Metta, G., & Wykowska, A. (2017). Robots as intentional agents: Using neuroscientific methods to make robots appear more social. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01663

Wiese, E., Weis, P. P., & Lofaro, D. M. (2018). Embodied social robots trigger gaze following in real-time HRI. *Proceedings of the 2018 International Conference on Ubiquitous Robots*. https://doi.org/10.1109/urai.2018.8441825

Yanco, H. A., & Drury, J. L. (2002). A taxonomy for human-robot interaction. *Proceedings of the AAAI fall symposium on human-robot interaction*, 111–119.

You, H.-C., & Lin, K.-W. (2019). Gendered tour-guide robots and their influence on user attitude and behavior. *Proceedings of the 2019 International ACM In-Cooperation HCI and UX Conference*. https://doi.org/10.1145/3328243.3328248

Yu, C., Fu, C., Chen, R., & Tapus, A. (2022). First attempt of gender-free speech style transfer for genderless robot. *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, 1110–1113.

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, *2*(3), 1497–1517. https://doi.org/10.1111/j.1751-9004.2008.00109.x

Złotowski, J., Khalil, A., & Abdallah, S. (2020). One robot doesn't fit all: Aligning social robot appearance and job suitability from a middle eastern perspective. *AI &SOCIETY*, *35*(2), 485–500. https://doi.org/10.1007/s00146-019-00895-x

Złotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2014). Anthropomorphism: Opportunities and challenges in human–robot interaction. *Inter-*

national Journal of Social Robotics, *7*(3), 347–360. https://doi.org/10.1007/s12369-014-0267-6