Eileen Roesler, Dietrich Manzey, Linda Onnasch

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

Citation details
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.

The Effects of Anthropomorphism on Human-Robot Interaction: A Quantitative Meta-Analysis

E. Roesler[1]*, D. Manzey[1], & L. Onnasch[2]

[1] Technische Universität Berlin, Germany, [2] Humboldt-Universität zu Berlin, Germany

* Corresponding author (eileen.roesler@tu-berlin.de)

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).

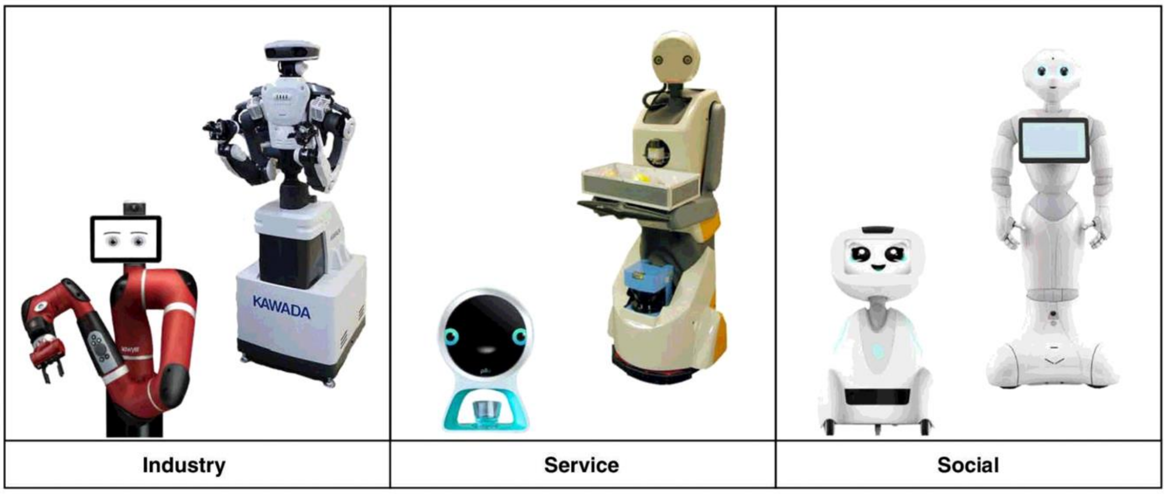

Fig. 1. Examples of anthropomorphic implementations. Anthropomorphic design by means of depicting human-like facial features or body features for the industrial (left: Sawyer; right: Nextage), service (left: Pillo Health ; right: SnackBot), and social domain (left: BUDDY; right: Pepper) received from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

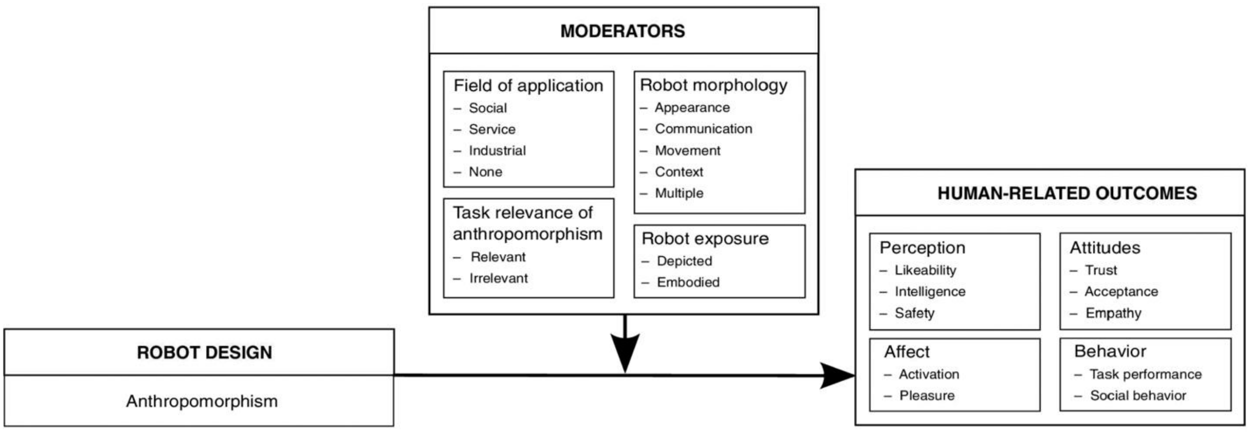

Fig.2. Basic framework of the meta-analysis.

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.

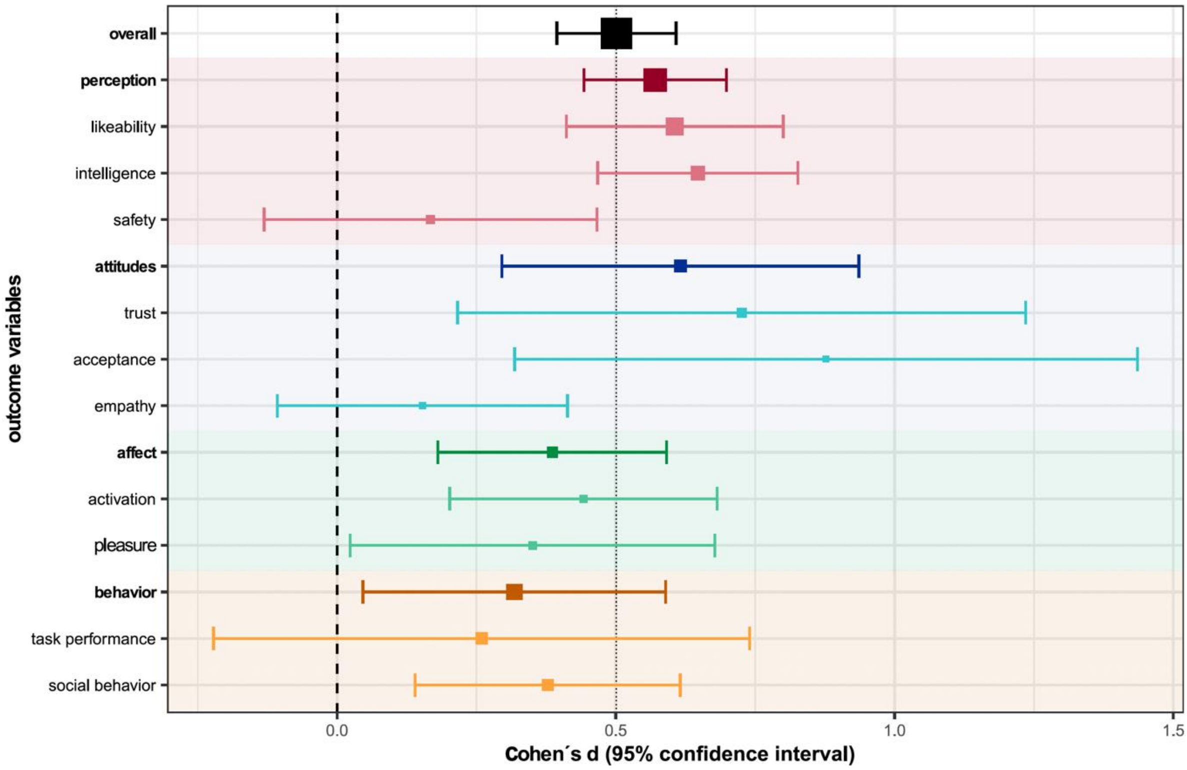

Fig. 3. Forest plot of the overall effect size and all sublevels. Depiction of standardized mean differences (Cohen's d) shown by the positions of the squares, the 95% CIs by the whiskers, and the numbers of included studies by the size of the squares.

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).

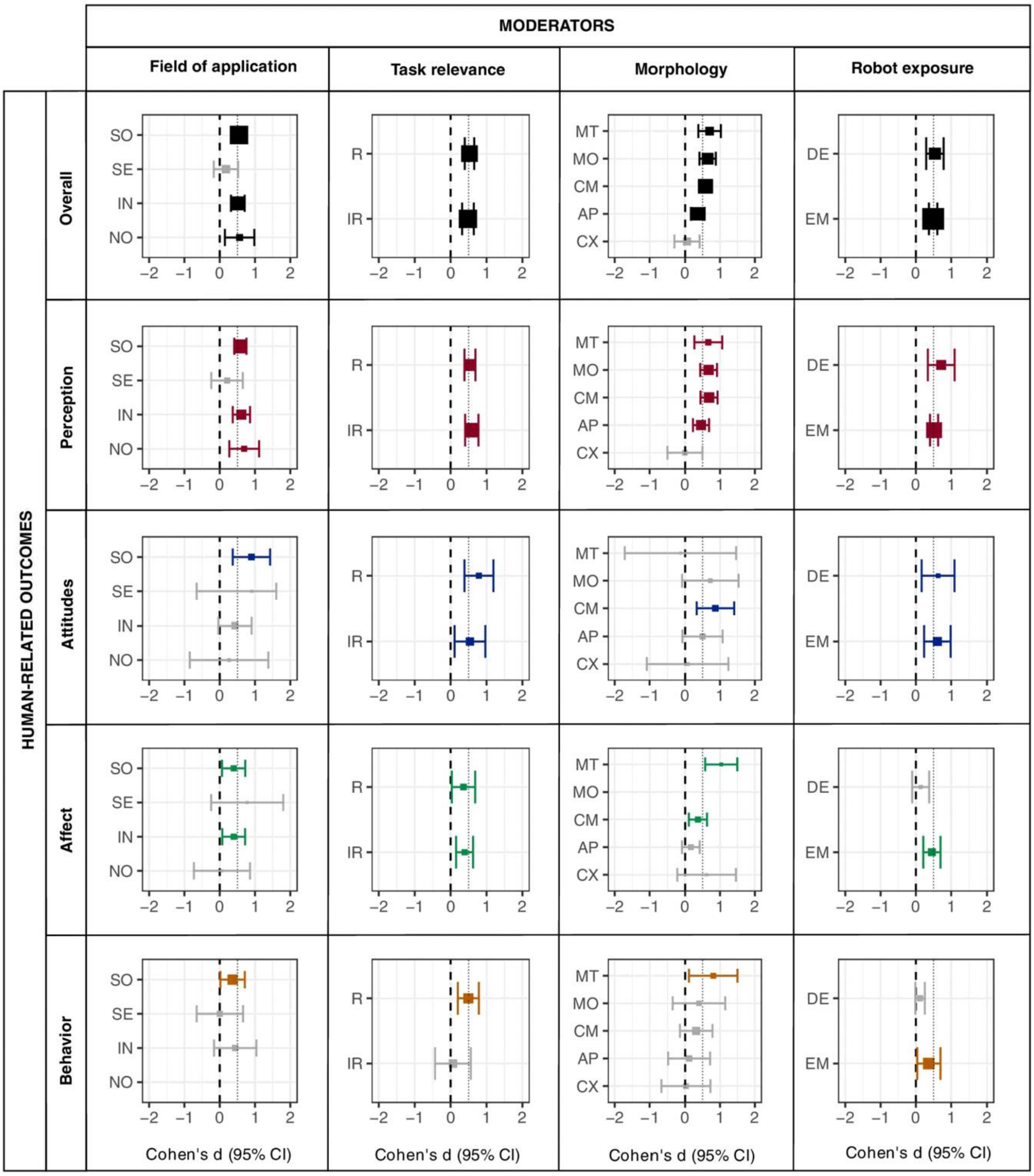

Fig. 4. Forest plots illustrating the effects of moderators. The plots show standardized mean differences (Cohen's d), the 95% CIs, and the number of effect sizes included, given separately for the overall effect and all subcategories, dependent on the characteristics of the different moderators (columns). 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.

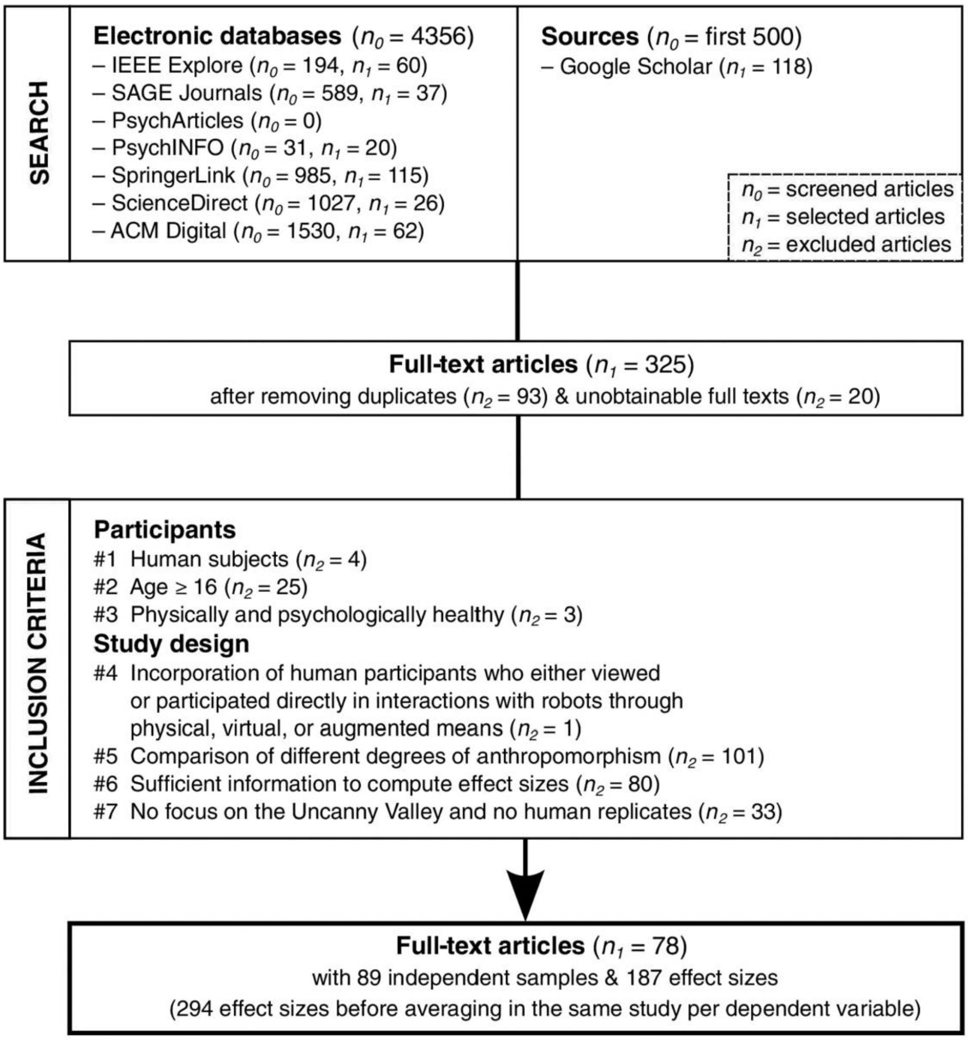

Fig. 5. Search flow diagram. Depiction of the entire process of data collection, including the sources searched, the inclusion criteria, and the selected articles.

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.