Virtually the Same or Realistically Different?: A Meta-analysis of Real vs. 'Not So Real' Robots

Connor Esterwood
School of Information
University of Michigan
Ann Arbor, USA

Ruijian (Hannah) Guan
School of Information
University of Michigan
Ann Arbor, USA

Xin Ye
School of Information
University of Michigan
Ann Arbor, USA

Lionel P. Robert
School of Information
University of Michigan
Ann Arbor, USA

***Abstract*—This study examined an important debate in Human-Robot Interaction (HRI) research: the suitability of non-physically non-collocated robots instead of physically collocated robots for HRI research. This meta-analysis (N=34 studies) examined the equivalence of physically and non-physically collocated robots in HRI research, focusing on anthropomorphism, social presence, and user engagement. No significant differences were found, suggesting that non-physical representations are viable alternatives. However, observed heterogeneity indicates potential moderating factors (e.g., task complexity, user characteristics, design features) warranting further investigation. These findings inform choices in resource-constrained environments.**

***Index Terms*—Human-Robot Interaction, Meta-Analysis, Robot Type, Embodiment**

I. Introduction

In the human–robot interaction (HRI) field, many studies have utilized non-physically collocated robots rather than physically collocated robots to explore research questions [1], [2]. These non-physically collocated robots take many forms based on the type and nature of the media used, including computer simulations, video recordings, photographs, and digital models of robots represented via virtual reality [3]–[5], on-screen simulations [6]–[8], and even videos or printed images [9], [10]. These non-physical representations provide a more economical, convenient, and controlled way to gain insights into how people perceive and respond to robots [11]–[13]. This is particularly important in countries and institutions that lack the resources to maintain physical robots.

However, a common and ongoing debate in the field concerns the validity of findings from studies that rely on non-physically collocated robots [14]–[18]. This concern is frequently highlighted in the limitations sections of these studies, as researchers question whether the results derived from non-physical representations of robots accurately reflect the dynamics of humans' in-person interactions with physically collocated robots. More specifically, there is an ongoing debate on whether robots non-physically collocated robots can elicit the same level of social presence, anthropomorphism, and engagement as physically collocated robots [19]–[22], all key drivers of trust, acceptance, and attitudes towards robots [21], [23]–[33]. Therefore, this issue poses a significant challenge to the HRI field, because it not only calls into question the validity of existing results and deployed measures but also shapes future research.

Existing research comparing physically and non-physically collocated robots yields inconsistent findings regarding the impact of embodiment on HRI outcomes [19]–[22]. This inconsistency hinders a clear understanding of the influence of robot interaction modality. Further research is needed to determine the overall effect of embodiment on HRI outcomes.

This study contributes to the ongoing debate regarding the impact of physical collocation in HRI. Our findings suggest that the discourse should shift from whether differences exist between physically and non-physically collocated robots to identifying the contexts in which these differences may become significant. Specifically, this study challenges previous qualitative findings suggesting superior social presence with physical robots [19], [20]. It also partially corroborates existing meta-analyses on anthropomorphism while highlighting inconsistencies regarding objective outcomes of engagement [21], [22]. Finally, it underscores the need for contextualized research focusing on functional embodiment and holistic design approaches.

II. Background

A. Embodiment Hypothesis

The Embodiment Hypothesis, rooted in embodied cognition, originates in the work of philosophers like Maurice Merleau-Ponty and George Lakoff, who emphasized the role of lived experience in shaping cognition [34], [35]. Embodied cognition posits that cognitive functions are deeply integrated with physical experiences [36], challenging traditional dualistic views of the mind and body [37]. This perspective suggests that cognition is not solely a product of internal representations but is shaped and influenced by embodied experiences.

Although the Embodiment Hypothesis has gained considerable support, it has also faced criticism. Some argue that it overemphasizes the role of the body and underestimates the importance of abstract, amodal representations in cognition [38]. Others contend that, while embodiment might play a role, it might not be as central or universal as proponents of the embodiment hypothesis claim [39]. Furthermore, scholars like Ziemke [40] distinguish between physical embodiment and functional embodiment. The latter emphasizes the role of a robot's behaviors and functional capabilities in shaping its perceived presence. This suggests that interactive function is as important as physical form.

Esterwood, C., Guan, R., Ye, X. and Robert, L. P. (2025). **Virtually the Same or Realistically Different?: A Meta-analysis of Real vs. 'Not So Real' Robots,** *Proceedings of the 20th ACM/IEEE International Conference on Human-Robot Interaction* (**HRI 2025**), (Acceptance rate: 25%) March 4-6, 2025, Melbourne, Australia.

B. Embodiment Hypothesis in HRI

The original Embodiment Hypothesis, built on the work of Merleau-Ponty and Lakoff, suggested that our bodily experience and physical interactions with the world fundamentally shape our perceptions, thoughts, and language [41]–[43]. In the HRI literature, the Embodiment Hypothesis is used to refer to the assertion that the robot's "physical embodiment can increase engagement and enjoyment in social interactions with humans" [20, Pg. 255]. Whereas scholars such as Merleau-Ponty and Lakoff assert that our perception of the world is rooted in our bodily experiences, the HRI literature asserts that our reaction to the other (e.g., robot) is influenced by that other's physical embodiment or lack thereof.

In this paper, we focus on the HRI-specific version of the Embodiment Hypothesis because this version drives the discussion around the need for a physically collocated robot. Indeed, if a robot's physical form significantly influences how humans perceive and interact with robots [44], [45], it may be that users interact with physically collocated and non-collocated robots in different ways [22], [32], [44].

HRI's Embodiment Hypothesis has generated considerable attention in HRI, including numerous meta-analyses and literature reviews [19]–[22]. These studies focused on three key HRI concepts: anthropomorphism, social presence, and engagement. These concepts are considered crucial in understanding how users perceive robots differently when presented non-physically, making them key factors for comparing the impact of non-physically collocated and physically collocated robots.

Anthropomorphism, defined as the attribution of human-like characteristics to non-human agents [46], has been shown to influence performance, acceptance, attitudes, and trust toward robots [21], [32], [33]. Social presence, characterized by the experience of non-physical or artificial agents as real social actors [47], has been linked to trust, acceptance, and positive attitudes toward technology in general [48]–[50] and robots in particular [29]–[31]. Finally, engagement, defined as the user's intent to maintain a connection with a robotic agent while performing a task [51], has been associated with performance, satisfaction, enjoyment, and acceptance of robots [23]–[28].

Figure 1 highlights how the relationship between physical embodiment and user perceptions of robots has been studied through qualitative and quantitative methods. This paper diverges from previous research in several key aspects. Qualitative reviews have suggested that physical robots elicit higher social presence and engagement [19], [20]. In contrast, we employed a quantitative meta-analysis to examine these relationships. Previous quantitative meta-analyses have explored the downstream effects of anthropomorphism and the potential moderation of measurement type [21], [22]. One study found no significant differences in anthropomorphism effects between non-physically collocated robots and physically collocated robots [21], while another revealed that physical embodiment leads to higher outcomes for objectively measured variables but not for subjective measures [22]. Our study uniquely examined the upstream effects of robot interaction modality on anthropomorphism, social presence, and engagement as direct outcomes. By doing so, we contribute a comprehensive and nuanced understanding of the ongoing debate about the equivalence of non-physically collocated and physically collocated robots in human–robot interaction research. Figure 1 visually represents our study's position within the existing literature, illustrating the current landscape of reviews and meta-analyses. As such, the paper's research question is as follows:

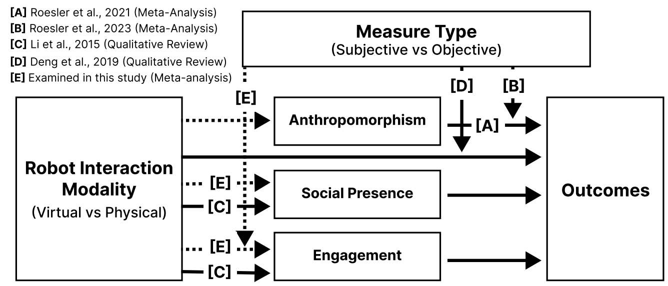

Fig. 1: This figure illustrates the current state of research. Arrows indicate relationships between variables that have been explored in meta-analyses or reviews. Arrows without references suggest that these relationships have not been examined.

> ***RQ:*** *To what extent do subjects' perceptions of anthropomorphism, social presence, and engagement differ between non-physically collocated and physically collocated robots?*

Addressing this research question is crucial because it has far-reaching implications for the validity of HRI research. If non-physically collocated robots yield fundamentally different results compared to physically collocated robots, a significant portion of existing HRI knowledge may be called into question. This concern is particularly pressing given the prevalence of online studies using non-physically collocated robots and their role in validating many subjective measurement instruments that are widely used in the field [21]. Consequently, the potential invalidity extends beyond studies employing non-physically collocated robots to those utilizing measures developed with such representations, even when conducted with physically collocated robots. This situation underscores the critical need to evaluate the comparability between non-physically collocated and physically collocated robots comprehensively.

The subsequent sections detail our methodology, analysis, and results, outlining the procedures for literature identification and meta-analytic techniques. Comprehensive methodological details, including robustness tests, are provided in the Appendix. We then discuss the implications of our findings in the context of existing literature, followed by an examination of study limitations. The paper concludes by synthesizing key insights and their significance for human–robot interaction.

III. METHOD

For this study we used Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA)-compliant systematic review and meta-analysis to compare responses to physically and non-physically collocated robots. Our methodology follows established HRI meta-analytic approaches [52]. This methodology mirrors the approaches of other HRI meta-analyses [53]–[56]. All data and code used for this analysis are visible at: 10.5281/zenodo.14270756.

A. Inclusion & Exclusion Criteria

This study employed a multi-level screening approach where criteria were progressively tightened at each stage. At the highest level (level-1 inclusion criteria), publications were deemed eligible for inclusion if they were classified as academic works (peer-reviewed publications, theses, dissertations, etc.), written in English, their titles or abstracts contained one or more of our search terms, they appeared empirical, and they included interactions between at least one human and at least one physically collocated or non-physically collocated robot. At the second level (level-2 inclusion criteria), publications were deemed eligible if they met all prior screening criteria, focused on non-physically collocated and physically collocated embodied physical action (EPA) robots, and included interactions between at least one human and at least one physically collocated and non-physically collocated robot. At the third level (level-3 inclusion criteria), publications were deemed eligible for inclusion if they met all prior screening criteria and directly compared outcomes of interactions with physically collocated robots *and* those with non-physically collocated robots. In cases where uncertainty existed at any level of inclusion, publications were deemed eligible and subjected to the subsequent level of screening.

In addition to these inclusion criteria, a set of exclusion criteria was applied. These exclusion criteria were applied at all levels. Specifically, publications were excluded if they focused on telepresence robots, did not compare physically collocated and non-physically collocated robots on the same outcome, utilized drastically different robots in the non-physical condition from those in the physical condition, or did not contain any direct interaction between a human and robot.

B. Data Sources and Database Search

This systematic review and meta-analysis leveraged Google Scholar, IEEE Explore, Scopus, and the ACM Digital Library. Our search took place between May 20 and June 1 of 2024. The terms used were developed through examination of keywords and consultation with a subject specialist librarian. The keywords used were: virtual, simulated, remote, simulator, disembodied, computer simulation, video recording, real-world, real life, physical, physical world, collocated, embodied, compare, comparison, contrast, differences, human–robot interaction, HRI, human–robot collaboration, and social robot.

Our search returned 73,558 results before accounting for duplicates, with 73,200 resulting from Google Scholar alone. In light of this, we adopted a pre-screening process for Google Scholar because its search algorithm provides a significantly broader range of results than the other databases. This process consisted of systematically examining each database's results on a page-by-page basis until no page provided a single result that met our level-1 inclusion criteria. For Google Scholar, the first 24 pages (10 results per page) passed this criterion, with the 25th page of results failing to contain a single result that passed our level-1 inclusion criteria. After these processes, a total of 805 total results across databases were extracted, de-duplicated, and exposed to title and abstract screening. These results were then extracted for screening via native `.BIB` or `.RIS` exporters in the case of IEEE Explore, Scopus, and the ACM Digital Library and via the Publish or Perish application [57] for Google Scholar. Results were then compiled in RAYYAN for further processing [58].

C. Study Selection and Screening

De-duplication, title, and abstract screening took place in RAYYAN [58]. RAYYAN is a free web and mobile app that helps expedite the initial screening of abstracts and titles and has been leveraged in many systematic reviews and meta-analyses, as evidenced by the software's 13,000 citations. Leveraging RAYYAN, we de-duplicated our search results in a semi-automatic fashion, creating a list of likely duplicate records and then manually screening them. This de-duplication process removed 173 records, leaving 632.

Title and abstract screening was performed in two phases, with level-1 inclusion criteria resulting in the exclusion of 289 records and level-2 inclusion criteria resulting in the exclusion of an additional 259 records. After this, a full-text screening was conducted based on our level-3 inclusion criteria. This screening incorporated the 84 records resulting from prior screening and an additional 58 records identified in prior reviews [19], [20]. This process led to 77 relevant publications emerging from the review. Figure 2 summarizes this screening process.

D. Categorization of Outcomes

The various studies examined a wide range of outcomes, with most investigations assessing multiple outcomes. To categorize and streamline these findings, the research team conducted an open card sort to distill specific outcomes into broader constructs, or "bins." Each team member grouped different outcomes and discussed these classifications collectively until a consensus was reached on the suitability of the resulting groups. Ultimately, 16 distinct outcome bins were established. [1] Among these, anthropomorphism, social presence, and engagement outcomes were selected for detailed analysis in this study given their direct relevance to the Embodiment Hypothesis as conceptualized and examined in HRI. Of the included studies, 12 provided sufficient effect size data for anthropomorphism, 12 for social presence, and 10 for engagement. We then classified the type of measure used

[1] A detailed list of these outcome bins is presented in the appendix

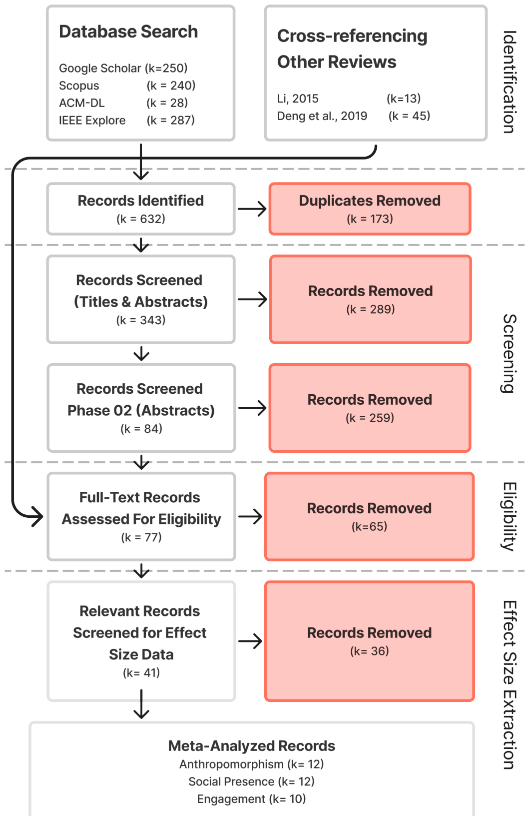

Fig. 2: Prisma diagram summarizing study screening and inclusion procedures.

for each of these outcomes as either subjective or objective. Subjective measures are psychological assessments that require judgment or interpretation and typically take the form of self-reports, while objective measures are directly observable behaviors such as gaze and interaction time [59]–[61]. Across these outcomes, only engagement utilized objective measures.

E. Data & Effect Size Extraction

Data from each study were extracted manually from each publication. This data included each paper's title, abstract, sample size, independent variables, dependent variables, and the reliabilities (α). In addition, we logged whether the non-physically collocated robots were presented to subjects on a 2-D screen or via virtual reality (VR) technologies. In all cases, these data were entered into a PostgreSQL-based database and exported to a `.xlsx` spreadsheet.

Using this spreadsheet, we then calculated effect sizes (r) based on the reported statistical information present for each study. These effect sizes reflect the magnitude of a treatment effect or the relative strength of a relationship [62]. In this meta-analysis, we utilized Pearson's (r) values. This was the case because Pearson's r is more frequently reported, easier to interpret, and is generally more popular as an effect size measure in meta-analyses [53], [56], [63], [64]. Pearson's r was obtained directly from the selected publications where possible. However, when this was impossible, effect sizes were calculated using [65] or [66] based on the information reported in the associated publication. The appendix of this article shows additional details on how r was derived for each study.

F. Statistical Methods

1) Meta-Analytical Approach: The meta-analyses conducted in this paper were done in the psychometric meta-analysis tradition [67]. The psychometric approach to meta-analysis involves correcting the distributions of observed correlation coefficients to estimate the distribution of population correlation coefficients [62]. As a result, the estimation of effect sizes in this analysis was corrected based on sample size *and* measurement error (Cronbach's α), producing an adjusted n that was used in analysis [67]. This differs from other meta-analytical approaches as it produces a corrected effect size that is generally more reliable [68], [69]. Using these *corrected* effect sizes, we then calculated an average effect size ($\bar{r}$) for each outcome.

2) Heterogeneity: Heterogeneity in a meta-analysis represents the degree of variance present between effect sizes [62], [70]. Assessing heterogeneity is vital for meta-analyses because heterogeneity ultimately determines how reliable an average effect size is and how one should interpret it [70]. In cases where heterogeneity is high, it is likely that variance among effects is caused by more than measurement or random error alone [70]. This lowers the reliability of the average effect observed because its utility for explaining variance is reduced and therefore unaccounted for moderators may be present [62], [70]–[72]. In this paper we considered both I^2 and Q statistics as our primary metrics of heterogeneity. Additional details on heterogeneity and its calculations are available in the appendix.

3) Publication Bias: Publication bias is the degree to which "the research that appears in the published literature is systematically unrepresentative of the population of completed studies" [73, Pg.1]. We constructed and statistically assessed funnel plots to determine the publication bias present across the meta-analyses examined in this paper. Additional details on funnel plots and how they are assessed are in the appendix.

4) Sensitivity & Outlier Detection: It was also important to ensure that our meta-analysis did not contain any studies that were outliers. To assess this, we conducted a leave-one-out sensitivity analysis. This analysis assists in identifying outliers by determining the impact that the exclusion of any single study has on overall findings. This is accomplished by running separate meta-analyses (one per study) excluding a different study each time. Results can then be inspected manually or via the `check.outliers()` function in r [74] to determine whether any of the analyses in our meta-analysis fell outside the 95% confidence interval of the total analysis of all other studies. In cases where this occurs, a study was considered an outlier and excluded from subsequent analysis [71], [74].

IV. Study & Sample Characteristics

For the outcome of anthropomorphism, the average sample size of studies was n=63 (SD=34). Within this sample, 56% of subjects on average were women indicating a relatively even distribution between male and female subjects [2]. The average age of subjects was 29 years (SD=17). The majority of studies examining anthropomorphism used screen-based non-physical representations (k=7) about as often as they used virtual reality representations of robots (k=6). Anthropomorphism was measured across studies exclusively via subjective measures with the majority using sub-dimensions or adaptations of the Godspeed questionnaire (k=5) [75] or custom measures (k=5).

For the outcome of social presence, the average sample size of studies was n=50 (SD=36). Within this sample, 46% of subjects on average were women, indicating a relatively even distribution between men and women. The average age of subjects was 31 years (SD=20). The majority of studies examining social presence used screen-based non-physical representations (k=7) as opposed to virtual reality representations of robots (k=4). Social presence was measured across studies via subjective measures in all but one case [76]. Of the subjective measures used, no single measure appeared dominant, with each study leveraging a related but different measure of social presence.

For the outcome of engagement, the average sample size of studies was n=41 (SD=32). Within this sample, 50% of subjects on average were women, indicating a relatively even distribution between male and female subjects. The average age of subjects was 28 years (SD=25). The majority of studies examining engagement used screen-based non-physical representations (k=7) as opposed to virtual reality representations of robots (k=2). Engagement was measured across studies via subjective measures (k=5) and objective (observational) measures (k=6) cases. No subjective or objective measure emerged consistently across these studies, with each study using a unique measure of engagement. A detailed breakdown of these measures and other study characteristics is presented in the appendix.

V. Meta-Analytic Results

A. Anthropomorphism

The overall corrected effect of robot interaction modality on subjects' perceptions of anthropomorphism indicated a non-significant overall effect (k=11, $\bar{r^2}$=-0.01, 95% CI;[-0.18,0.15]). This implies that robot interaction modality does not significantly impact subjects' perceptions of anthropomorphism. This analysis was seen as relatively robust given minimal publication bias (Egger's $t = 0.32$, $df = 9$, $P = 0.76$) and relatively stable results in leave-one-out sensitivity analysis after the removal of outliers[3]. Significant heterogeneity, however, was observed ($Q(df = 10) = 28.11$, $p - val = 0.002$; $I^2 = 64\%$)), indicating that a large portion of variance among the observed effect sizes may be due to sources other than random error. As a result, it is likely that moderators may be at play and additional analysis is needed [62], [70]. These results are summarized in row 1 of table I and are visually presented in Figure 3.

Meta-Analysis of Anthropomorphism with Individual Correction								Heterogeneity	
Analysis	*k*	*N*	$\bar{r}$	*Var r*	*CI LL*	*CI UL*	*Sig*	I^2	*Q*
Non-physical (Both)	11	642	-0.015	0.049	-0.178	0.151	N	64%	P=0.11
Screen	7	501	-0.07	0.025	-0.231	0.084	N	42.37%	P=0.49
VR	6	326	-0.025	0.091	-0.381	0.339	N	79%	P $<$ 0.001

TABLE I: Meta-analytical results for anthropomorphism including sub-group analysis by non-physical representation.

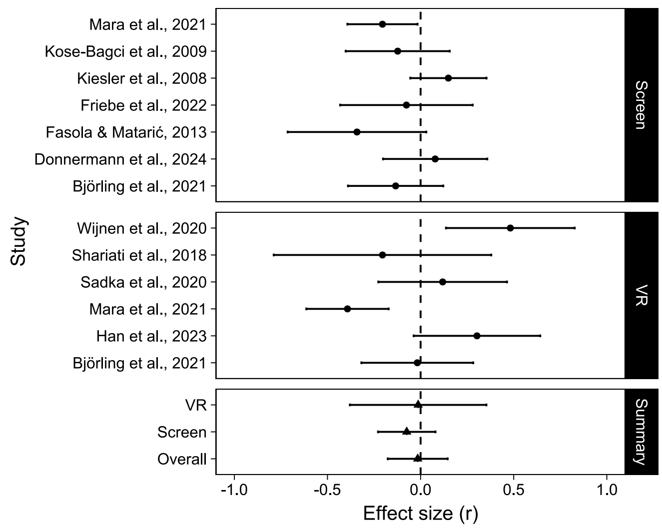

Fig. 3: Forest plot for anthropomorphism illustrating corrected effect sizes for the impact of robot interaction modality by type of non-physical representation. Each dot represents the effect size of a study, and the line around it shows the uncertainty in that estimate. The overall effect size is shown at the bottom as a composite of the effect sizes above.

1) Sub-Group Analysis: Given the degree of heterogeneity detected in our overarching meta-analysis, we conducted a follow-up sub-group analysis. We did this to determine whether the variance observed between effect sizes (i.e., heterogeneity) might be attributable to one or more moderating factors. In particular, we hypothesized that the type of non-physical representation used (screen-based or VR) may determine when said non-physical representation is more or less influential. Results of sub-group analysis, however, indicated that this was unlikely. In particular, neither the sub-group analysis for screens alone (k=7, $\bar{r^2}$=-0.07, 95% CI;[-0.23,0.08]) nor analysis for virtual reality (k=6, $\bar{r^2}$=-0.03, 95% CI;[-0.38,0.34]) observed effects for robot interaction modality. It is worth noting, however, that the confidence intervals for the sub-group analysis on screens (CI$[-0.23, 0.08]$) did appear relatively narrow and that heterogeneity in this group ($Q(df = 6) = 10.41$, $p - val = 0.11$; $I^2 = 42.37\%$) was reduced to a point where additional moderators were unlikely to manifest. This indicates that in the case of screens, some

[2] Studies did not provide data on the number of non-binary subjects

[3] [77] was detected as an outlier and removed from this analysis. Table I in the appendix shows the sensitivity analysis where said outlier was detected and an analysis including this outlier is presented in this paper's appendix. Results with outliers included did not differ significantly from the results shown here.

effect on anthropomorphism may be present but that said effect may merely be too small to register with the current sample (k = 7 studies). Table I summarizes these findings in rows 2 and 3, while Figure 3 illustrates these effects visually.

B. Social Presence

The overall corrected effect of robot interaction modality on subjects' perceptions of social presence indicated a non-significant overall effect (k=9, $\bar{r^2}$=-0.12, 95% CI;[-0.41,0.16]). This implies that robot interaction modality does not have an impact on the degree to which humans see said robots as possessing social presence. This analysis was seen as relatively robust given minimal publication bias (Egger's $t = 1.09$, $df = 7$, $P = 0.31$) and relatively stable results in leave-one-out sensitivity analysis after the removal of outliers. [4] It is important to note, however, that sizable heterogeneity was visible across this overall effect ($Q(df = 8) = 3.87$, $p - val > 0.001$; $I^2 = 79\%$), indicating that a large portion of variance among the observed effect sizes may be due to sources other than random error. As a result, it is likely that moderators may be at play and additional analysis is needed [62], [70]. These results are summarized in row 1 of table II and are visually presented in Figure 4.

Meta-Analysis of Social Presence With Individual Correction								Heterogeneity	
Analysis	*k*	*N*	$\bar{r}$	*Var r*	*CI LL*	*CI UL*	*Sig*	I^2	*Q*
Non-Physical (Both)	9	376	-0.12	0.12	-0.41	0.16	**N**	79%	P<0.001
Screen	7	302	-0.25	0.1	-0.57	0.04	N	77%	P<0.001
VR	4	259	-0.001	0.06	-0.45	0.45	N	74%	P=0.007

TABLE II: Meta-analytical results for social presence including sub-group analysis by non-physical representation type.

1) Sub-Group Analysis: Given the degree of heterogeneity detected in our overarching meta-analysis, we conducted a follow-up sub-group analysis. We did this to determine whether the variance observed between effect sizes (i.e., heterogeneity) might be attributable to one or more moderating factors. In particular, we hypothesized that the type of non-physical presentation used (screen-based or VR) may determine when the impact of using a non-physical representation is more or less influential in terms of social presence. Similar to our results for anthropomorphism, however, these results indicated that this was unlikely. In particular, neither the sub-group analysis for screens alone (k=7, $\bar{r^2}$=-0.25, 95% CI;[-0.57,0.04]) nor that for virtual reality (k=4, $\bar{r^2}$=-0.001, 95% CI;[-0.45,0.45]) observed effects for robot interaction modality. This, in combination with no sizable shifts in heterogeneity statistics – $Q(df = 6) = 2.7$, $p - val < 0.001$; $I^2 = 77.64\%$ and $Q(df = 3) = 11.88$, $p - val = 0.008$; $I^2 = 74.75\%$ respectively – indicates that the type of non-physical representation (VR vs. screen) is unlikely to impact the overall effect of using non-physical representations of robots as opposed to physically collocated robots with regard to social presence. Table II summarizes these findings in rows 2 and 3 while Figure 4 illustrates these effects visually.

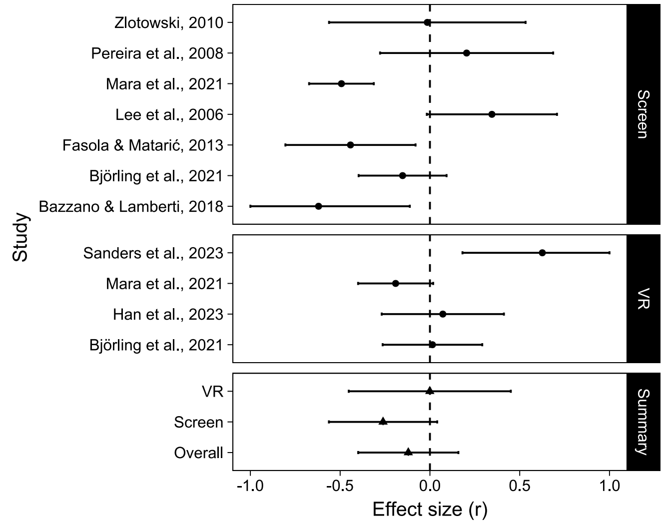

Fig. 4: Forest plot for social presence illustrating corrected effect sizes for the impact of robot interaction modality by type of non-physical representation.

Meta-Analysis of Engagement With Individual Correction								Heterogeneity	
Analysis	*k*	*N*	$\bar{r}$	*Var r*	*CI LL*	*CI UL*	*Sig*	I^2	*Q*
Non-Physical (Both)	9	402	-0.19	0.09	-0.41	0.04	**N**	75%	P<0.001
Objective Measures	5	318	-0.17	0.12	-0.61	0.26	**N**	88%	P<0.001
Subjective Measures	5	146	-0.08	0.05	-0.39	0.22	**N**	22%	P=0.27

TABLE III: Meta-analytical results for engagement, including sub-group analysis based on the type of engagement measure.

C. Engagement

The overall corrected effect of robot interaction modality on subjects' engagement indicated a non-significant overall effect (k=9, $\bar{r^2}$=-0.19, 95% CI;[-0.42,0.04). This implies that robot interaction modality does not have an impact on the degree to which humans engage with said robots. This analysis was seen as relatively robust given minimal publication bias (Egger's $t = -0.56$, $df = 7$, $P = 0.59$) and relatively stable results in leave-one-out sensitivity analysis after the removal of outliers[5]. It is important to note, however, that sizable heterogeneity is visible across this overall effect ($Q(df = 8) = 3.2$, $p - val > 0.001$; $I^2 = 75\%$), indicating that a large portion of variance among the observed effect sizes may be due to sources other than random error. As a result, it is likely that moderators may be at play and additional analysis is needed [62], [70]. These results are summarized in Table III and are visually presented in Figure 5.

1) Sub-Group Analysis: Given the heterogeneity detected in our overarching meta-analysis, we sought to conduct a follow-up sub-group analysis. Due to the small number of studies

[4] [16], [31], [78] were detected as outliers and removed from this analysis. The appendix shows the sensitivity analysis where said outliers were detected and an analysis including these outliers is available in the appendix. Results with outliers did not differ significantly from the results shown here

[5] [79] was detected as an outlier and removed from this analysis. Table III in the appendix shows the sensitivity analysis where said outlier was detected and an analysis including this outlier is visible in this paper's appendix. Results with outliers included do not differ significantly from the results shown here.

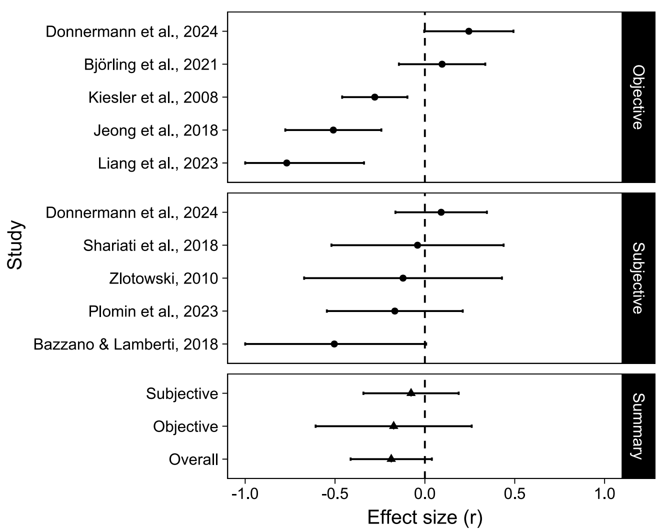

Fig. 5: Forest plot for engagement illustrating corrected effect sizes for the impact of robot interaction modality divided by type of engagement measure.

that used virtual reality as their non-physical representation of a robot ($k = 2$), however, such sub-group analysis was not possible. That said, engagement was unique from our other outcomes in that this outcome was examined with both subjective and objective measures. Given that the type of measurement is influential for other outcomes [21], we therefore explored the possibility of measurement type acting as an alternative moderator. In particular, we hypothesized that the kind of outcome measure used (objective or subjective) may determine when the impact of using a non-physical representation is more or less influential.

Results of this sub-group analysis, however, indicated that this was unlikely. In particular, neither the sub-group analysis for objective measures (k=5, $\bar{r^2}$=-0.17, 95% CI;[-0.61,0.26]) nor that for subjective measures (k=5, $\bar{r^2}$=-0.08, 95% CI;[-0.39,0.22]) observed effects for robot interaction modality. Notably, heterogeneity was reduced in the sub-group using subjective measures ($Q(df = 4) = 0.25$, $p - val = 0.27$; $I^2 = 23\%$), however, the lack of a similar trends in the case of objective measures ($Q(df = 4) = 32$, $p - val < 0.001$; $I^2 = 88\%$), as well as non-significant overall effects imply that this reduction is likely due to other reasons. These findings indicate that the type of measure used is unlikely to impact the overall effect of using non-physical representations of robots as opposed to physically collocated robots with regard to engagement. Table III summarizes these findings in rows 2 and 3 while Figure 5 illustrates these effects visually.

VI. Summary of Results

Overall, our results indicate that robot interaction modality does not significantly impact humans' perceptions of a robot's anthropomorphism or social presence. In addition, it does not appear that robot interaction modality significantly affects humans' engagement with robots. That said, each meta-analysis conducted contained a strong degree of heterogeneity (i.e., variance among effect sizes) that was not accounted for. This indicates that moderators are likely present and may influence the effect of robot interaction modality on each of these outcomes.

This study explored the non-physical representation type (virtual reality vs. screen) as one such moderator. Subgroup analysis, however, revealed that for social presence and anthropomorphism, the method of non-physical representation was not a moderator; as for engagement, insufficient studies employing virtual reality precluded a meaningful comparison. In addition to non-physical representation type, we explored the possibility that measurement or outcome type was a moderator. This was only possible for the outcome of engagement given that neither anthropomorphism nor social presence studies utilized objective measures. Results of this sub-group analysis were again non-significant, indicating that – at least in the case of engagement – the type of measure did not appear to moderate the impact of robot interaction modality.

VII. Discussion

This study challenges the assumption underlying the Embodiment Hypothesis, prompting its re-evaluation in HRI. This paper directly contradicts prior qualitative reviews suggesting that physical robots elicit higher social presence and engagement [19], [20]. In doing so, our findings indicate that the perceived advantages of a physical embodiment may be less pronounced or consistent than previously thought. However, it is important to note that the prior studies [19] and [20] were qualitative analyses of the literature as opposed to the quantitative analyses conducted in this paper. As a result, [19] and [20] did not examine statistical data directly. Although their approach was valid, this alone could explain the differences.

Our paper also corroborates and contradicts prior meta-analyses on the impact of robot interaction modality in HRI [21], [22]. On the one hand, our study aligns with previous findings because it did not observe significant differences between modalities for anthropomorphism or for subjective outcomes more broadly. On the other hand, however, this paper's results contradict Roesler et al. [22] because differences between non-physically collocated and physically collocated robots were observed for objective outcomes in [22] but were not in our meta-analysis with regard to engagement.

Scholars like Dove [38] and Goldinger et al. [39] argued that the role of physical embodiment might be overstated and that abstract, amodal representations can facilitate effective interaction. Our paper's results also support Ziemke's [40] distinction between physical and functional embodiment, where the functional capabilities of robots, regardless of physicality, can drive social presence, emphasizing the importance of behavioral and interactive design over merely physical form. It is essential to note that these findings do not directly speak to Maurice Merleau-Ponty and George Lakoff's version of the Embodiment Hypothesis. Interestingly, Merleau-Ponty and Lakoff's Embodiment Hypothesis may be invaluable to understanding how physical robots interact with the world in comparison to chatbots or other non-physical robots. As such, future work is needed to examine this version of the Embodiment Hypothesis more directly.

A. Implications for Theory and Future Research

Our results suggest a paradigm shift is necessary for HRI research. Rather than focusing on whether differences exist be- tween physically collocated and non-physically collocated robots, future investigations should identify the specific contexts and conditions under which these differences may become significant. This nuanced approach will provide a more comprehensive understanding of the role of embodiment in HRI. To address this, further studies, particularly those incorporating qualitative methodologies, are essential in the following areas.

First, it is crucial to investigate contextual factors that explain when there may or may not be differences between non-physical and physical robots. This includes examining the impact of different levels and types of embodiment (e.g., humanoid vs. non-humanoid, degree of anthropomorphism, etc.) [80] on key interaction outcomes, including anthropomorphism, social presence, and engagement.

Second, examining individual differences is essential for comprehending how personal characteristics, such as personality traits, technology experiences, and cultural backgrounds, influence user engagement with robots [81]. Such an understanding can lead to more tailored and effective interactions, providing valuable insights for the design and deployment of non-physically collocated and physically collocated robots across a wide range of applications.

Third, beyond physical embodiment, the concept of functional embodiment merits consideration. This perspective emphasizes the impact of a robot's behaviors, capabilities, and environmental interactions on human perceptions [40]. For example, a robot's task performance, communication efficacy, and responsiveness may be as influential as its physical form in shaping user experiences. However, further research is necessary to delineate the boundary conditions of this effect and its relative importance.

Fourth, future research should prioritize ecological validity to improve the generalizability of findings in real-world contexts. This focus extends beyond the simple comparison of physical versus non-physical robots and underscores the significance of contextual factors in shaping HRI. Researchers must consider interrelated elements such as task type and nature, user expectations, environmental context, and robot behavior and appearance to achieve high ecological validity. Understanding how these factors collectively influence user perceptions is essential for bridging the gap between laboratory findings and real-world applications.

Fifth, future studies should aim to develop an integrated theoretical framework that identifies the conditions under which non-physically collocated robot use is appropriate or inappropriate relative to the use of physically collocated robots. This approach should incorporate a deeper understanding of a task's nature, user's expectations, and environmental contexts. Addressing these research directions can advance our knowledge of the conditions under which non-physically collocated and physically collocated robot representations may yield divergent outcomes. This knowledge will provide crucial guidance on using non-physically collocated robots versus physically collocated robots across different applications and contexts within HRI research.

Finally, the results of this study also provide implications for design. Our findings, along with the existing literature, suggest a need for a holistic design approach that prioritizes behavioral capabilities and contextual responsiveness over physical presence. Designers should, therefore, focus on creating interactive and emotionally responsive functionalities and exploring innovative ways to replicate presence digitally. This can provide various benefits, including the more effective deployment of virtual agents in scenarios where physical robots are impractical due to logistical or cost constraints. This expands the range of applications for which non-physically collocated robots can be used, which increases their scalability.

VIII. Limitations and Future Work

As with any systematic review and meta-analysis, our current results reflect certain limitations within the literature. In particular, while sample diversity was high regarding gender, the average age of subjects was relatively young (under 35). As a result, these findings speak primarily to a younger population, and future work is required to explore whether these results shift with the inclusion of older populations.

The comparison between non-physically collocated and physically collocated robots often involves moderators and confounding variables, complicating the isolation of embodiment effects [20]. This study addressed these variables in two ways. First, by conducting a meta-analysis, it overcame the limitations of individual HRI studies, including confounders or moderators, by consolidating findings across studies [62], [70].Second, it focused exclusively on studies that directly compared non-physically collocated and physically collocated robots of the same type within the same study, creating a unique dataset that avoids issues from cross-study comparisons. Despite these efforts, unidentified confounders and moderators may have influenced the findings due to observed variance in effect sizes. As a result, the null result may be due to various issues other than embodiment alone. Future research should, therefore, explore possible confounders and moderators.

IX. Conclusion

These findings challenge traditional assumptions about physical embodiment in HRI, suggesting that robot designers should strategically leverage both non-physical and physical modalities based on application needs and user contexts. This research contributes to a more nuanced understanding of human-like presence in HRI, highlighting the complexity of achieving effective interaction beyond traditional boundaries.

References

[1] D. Feil-Seifer, K. S. Haring, S. Rossi, A. R. Wagner, and T. Williams, "Where to next? the impact of covid-19 on human-robot interaction research," pp. 1–7, 2020.

[2] N. Wang, D. V. Pynadath, K. Unnikrishnan, S. Shankar, and C. Merchant, "Intelligent agents for virtual simulation of human-robot interaction," in *Virtual, Augmented and Mixed Reality: 7th International Conference, VAMR 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings 7*. Springer, 2015, pp. 228–239.

[3] M. Duguleana, F. G. Barbuceanu, and G. Mogan, "Evaluating human-robot interaction during a manipulation experiment conducted in immersive virtual reality," in *Virtual and Mixed Reality-New Trends: International Conference, Virtual and Mixed Reality 2011, Held as Part of HCI International 2011, Orlando, FL, USA, July 9-14, 2011, Proceedings, Part I 4*. Springer, 2011, pp. 164–173.

[4] S. K. Jayaraman, C. Creech, D. M. Tilbury, X. J. Yang, A. K. Pradhan, K. M. Tsui, and L. P. Robert Jr, "Pedestrian trust in automated vehicles: Role of traffic signal and av driving behavior," *Frontiers in Robotics and AI*, vol. 6, p. 117, 2019.

[5] X. Ye and L. P. Robert, "Human security robot interaction and anthropomorphism: An examination of pepper, ramsee, and knightscope robots," in *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2023, pp. 982–987.

[6] S. Bhat, J. B. Lyons, C. Shi, and X. J. Yang, "Evaluating the impact of personalized value alignment in human-robot interaction: Insights into trust and team performance outcomes," in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 2024, pp. 32–41.

[7] H. Azevedo-Sa, X. J. Yang, L. P. Robert, and D. M. Tilbury, "A unified bi-directional model for natural and artificial trust in human–robot collaboration," *IEEE robotics and automation letters*, vol. 6, no. 3, pp. 5913–5920, 2021.

[8] C. Esterwood and L. P. Robert, "The theory of mind and human–robot trust repair," *Scientific Reports*, vol. 13, no. 1, p. 9877, 2023.

[9] Y.-C. Chang, D. J. Rea, and T. Kanda, "Investigating the impact of gender stereotypes in authority on avatar robots," in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 2024, pp. 106–115.

[10] X. Ye, W. Jo, A. Ali, S. C. Bhatti, C. Esterwood, H. A. Kassie, and L. P. Robert, "Autonomy acceptance model (aam): The role of autonomy and risk in security robot acceptance," in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 2024, pp. 840–849.

[11] T. Inamura and J. T. C. Tan, "Long-term large scale human-robot interaction platform through immersive vr system-development of robocup@ home simulator," in *2012 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 2012, pp. 242–247.

[12] S. Honig and T. Oron-Gilad, "Comparing laboratory user studies and video-enhanced web surveys for eliciting user gestures in human-robot interactions," in *Companion of the 2020 ACM/IEEE international conference on human-robot interaction*, 2020, pp. 248–250.

[13] Y. Lei, Z. Su, and C. Cheng, "Virtual reality in human-robot interaction: Challenges and benefits," *Electronic Research Archive*, vol. 31, no. 5, pp. 2374–2408, 2023.

[14] L. Wijnen, S. Lemaignan, and P. Bremner, "Towards using virtual reality for replicating hri studies," in *Companion of the 2020 ACM/IEEE international conference on human-robot interaction*, 2020, pp. 514–516.

[15] F. Babel, J. Kraus, P. Hock, H. Asenbauer, and M. Baumann, "Investigating the validity of online robot evaluations: Comparison of findings from an one-sample online and laboratory study," in *Companion of the 2021 ACM/IEEE international conference on human-robot interaction*, 2021, pp. 116–120.

[16] R. Li, M. van Almkerk, S. van Waveren, E. Carter, and I. Leite, "Comparing human-robot proxemics between virtual reality and the real world," in *2019 14th ACM/IEEE international conference on human-robot interaction (HRI)*. IEEE, 2019, pp. 431–439.

[17] N. Randall and S. Sabanovic, "A picture might be worth a thousand words, but it's not always enough to evaluate robots," in *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 2023, pp. 437–445.

[18] S. N. Woods, M. L. Walters, K. L. Koay, and K. Dautenhahn, "Methodological issues in hri: A comparison of live and video-based methods in robot to human approach direction trials," in *ROMAN 2006-the 15th IEEE international symposium on robot and human interactive communication*. IEEE, 2006, pp. 51–58.

[19] J. Li, "The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents," *International Journal of Human-Computer Studies*, vol. 77, pp. 23–37, 2015.

[20] E. Deng, B. Mutlu, M. J. Mataric *et al.*, "Embodiment in socially interactive robots," *Foundations and Trends® in Robotics*, vol. 7, no. 4, pp. 251–356, 2019.

[21] E. Roesler, D. Manzey, and L. Onnasch, "A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction," *Science Robotics*, vol. 6, no. 58, p. eabj5425, 2021.

[22] ——, "Embodiment matters in social hri research: Effectiveness of anthropomorphism on subjective and objective outcomes," *ACM Transactions on Human-Robot Interaction*, vol. 12, no. 1, pp. 1–9, 2023.

[23] A. Meissner, A. Trübswetter, A. S. Conti-Kufner, and J. Schmidtler, "Friend or foe? understanding assembly workers' acceptance of human-robot collaboration," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 10, no. 1, pp. 1–30, 2020.

[24] M. Paliga, "Human–cobot interaction fluency and cobot operators' job performance. the mediating role of work engagement: A survey," *Robotics and Autonomous Systems*, vol. 155, p. 104191, 2022.

[25] H. Tennent, S. Shen, and M. Jung, "Micbot: A peripheral robotic object to shape conversational dynamics and team performance," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2019, pp. 133–142.

[26] K. Drejing, S. Thill, and P. Hemeren, "Engagement: A traceable motivational concept in human-robot interaction," in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2015, pp. 956–961.

[27] S. Ng, T.-H. Lin, Y. Li, and S. Sebo, "Role-playing with robot characters: Increasing user engagement through narrative and gameplay agency," in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 2024, pp. 522–532.

[28] Y. Lin, W. Jo, A. Ali, L. P. Robert Jr, and D. M. Tilbury, "Toward personalized tour-guide robot: Adaptive content planner based on visitor's engagement," in *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 2024, pp. 674–678.

[29] N. Chen, X. Liu, and X. Hu, "Effects of robots' character and information disclosure on human–robot trust and the mediating role of social presence," *International Journal of Social Robotics*, vol. 16, no. 4, pp. 811–825, 2024.

[30] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, "Influence of social presence on acceptance of an assistive social robot and screen agent by elderly users," *Advanced Robotics*, vol. 23, no. 14, pp. 1909–1923, 2009.

[31] Y. Jung and K. M. Lee, "Effects of physical embodiment on social presence of social robots," *Proceedings of PRESENCE*, vol. 2004, pp. 80–87, 2004.

[32] J. Złotowski, D. Proudfoot, K. Yogeeswaran, and C. Bartneck, "Anthropomorphism: opportunities and challenges in human–robot interaction," *International journal of social robotics*, vol. 7, pp. 347–360, 2015.

[33] S. C. Bhatti and L. P. Robert, "What does it mean to anthropomorphize robots? food for thought for hri research," in *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 2023, pp. 422–425.

[34] G. Lakoff and M. Johnson, *Metaphors we live by*. University of Chicago press, 2008.

[35] K. Takaki, "Embodied knowing: The tacit dimension in johnson and lakoff, and merleau-ponty," *Tradition and discovery: the Polanyi society periodical*, vol. 36, no. 2, pp. 26–39, 2009.

[36] R. Pfeifer and C. Scheier, "Representation in natural and artificial agents: an embodied cognitive science perspective," *Zeitschrift für Naturforschung C*, vol. 53, no. 7-8, pp. 480–503, 1998.

[37] R. A. Brooks, "Intelligence without representation," *Artificial intelligence*, vol. 47, no. 1-3, pp. 139–159, 1991.

[38] G. Dove, "Three symbol ungrounding problems: Abstract concepts and the future of embodied cognition," *Psychonomic bulletin & review*, vol. 23, pp. 1109–1121, 2016.

[39] S. D. Goldinger, M. H. Papesh, A. S. Barnhart, W. A. Hansen, and M. C. Hout, "The poverty of embodied cognition," *Psychonomic bulletin & review*, vol. 23, pp. 959–978, 2016.

[40] T. Ziemke, "What's that thing called embodiment?" in *Proceedings of the 25th Annual Cognitive Science Society*. Psychology Press, 2013, pp. 1305–1310.

[41] M. Merleau-Ponty, "Phenomenology of perception," *Translated by Colin Smith*, 1965.

[42] L. George, "Women, fire, and dangerous things: What categories reveal about the mind," *Chicago: University of Chicago*, 1987.

[43] G. Lakoff and R. E. Núñez, "Where mathematics comes from: How the embodied mind brings mathematics into being." 2000.

[44] J. Wainer, D. J. Feil-Seifer, D. A. Shell, and M. J. Mataric, "The role of physical embodiment in human-robot interaction," in *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2006, pp. 117–122.

[45] N. Dennler, C. Ruan, J. Hadiwijoyo, B. Chen, S. Nikolaidis, and M. Matarić, "Design metaphors for understanding user expectations of socially interactive robot embodiments," *ACM Transactions on Human-Robot Interaction*, vol. 12, no. 2, pp. 1–41, 2023.

[46] N. Epley, A. Waytz, and J. T. Cacioppo, "On seeing human: A three-factor theory of anthropomorphism." *Psychological Review*, vol. 114, no. 4, pp. 864 – 886, 2007. [Online]. Available: http://proxy.lib.umich.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=pdh&AN=2007-13558-002&site=ehost-live&scope=site

[47] K. M. Lee, "Presence, explicated," *Communication theory*, vol. 14, no. 1, pp. 27–50, 2004.

[48] K. M. Lee and C. Nass, "Designing social presence of social actors in human computer interaction," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2003, pp. 289–296.

[49] C. DiSalvo, "All robots are not created equal: The design and perception of humanoid robot heads," *Human Computer Interaction Institute and school of Design, Carnegie Mellon University*, 2002.

[50] T. Bickmore and D. Schulman, "The comforting presence of relational agents," in *CHI'06 Extended Abstracts on Human Factors in Computing Systems*, 2006, pp. 550–555.

[51] A. Sorrentino, L. Fiorini, and F. Cavallo, "From the definition to the automatic assessment of engagement in human–robot interaction: A systematic review," *International Journal of Social Robotics*, pp. 1–23, 2024.

[52] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan *et al.*, "The prisma 2020 statement: an updated guideline for reporting systematic reviews," *bmj*, vol. 372, 2021.

[53] C. Esterwood, K. Essenmacher, H. Yang, F. Zeng, and L. P. Robert, "A meta-analysis of human personality and robot acceptance in human-robot interaction," in *Proceedings of the 2021 CHI conference on human factors in computing systems*, 2021, pp. 1–18.

[54] ——, "Birds of a feather flock together: but do humans and robots? a meta-analysis of human and robot personality matching," in *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, 2021, pp. 343–348.

[55] C. Esterwood and L. P. Robert, "Do you still trust me? human-robot trust repair strategies," in *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, 2021, pp. 183–188.

[56] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. De Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human factors*, vol. 53, no. 5, pp. 517–527, 2011.

[57] A. W. Harzing, "Publish or perish app," 2007. [Online]. Available: https://harzing.com/resources/publish-or-perish

[58] M. Ouzzani, H. Hammady, Z. Fedorowicz, and A. Elmagarmid, "Rayyan—a web and mobile app for systematic reviews," *Systematic Reviews*, vol. 5, no. 1, 2016.

[59] M. W. Schustack and H. S. Friedman, *Psychological Testing, Overview*. Elsevier, 2005.

[60] H. A. Ferreira and M. Saraiva, "Subjective and objective measures," *Emotional design in human-robot interaction: Theory, methods and applications*, pp. 143–159, 2019.

[61] J. A. Marvel, S. Bagchi, M. Zimmerman, and B. Antonishek, "Towards effective interface designs for collaborative hri in manufacturing: Metrics and measures," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 9, no. 4, pp. 1–55, 2020.

[62] M. Borenstein, L. V. Hedges, J. P. Higgins, and H. R. Rothstein, *Introduction to meta-analysis*. John Wiley & Sons, 2021.

[63] C. Esterwood, K. Essenmacher, H. Yang, F. Zeng, and L. P. Robert, "A personable robot: Meta-analysis of robot personality and human acceptance," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6918–6925, 2022.

[64] P. A. Hancock, T. T. Kessler, A. D. Kaplan, J. C. Brill, and J. L. Szalma, "Evolving trust in robots: specification through sequential and comparative meta-analyses," *Human factors*, vol. 63, no. 7, pp. 1196–1229, 2021.

[65] D. B. Wilson, "Practical meta-analysis effect size calculator," 2023.

[66] W. Lenhard and A. Lenhard, "Computation of effect sizes," 2022. [Online]. Available: https://www.psychometrica.de/effect_size.html

[67] J. A. Dahlke and B. M. Wiernik, "psychmeta: An r package for psychometric meta-analysis," *Applied psychological measurement*, vol. 43, no. 5, pp. 415–416, 2019.

[68] D. S. Ones, C. Viswesvaran, and F. L. Schmidt, "Realizing the full potential of psychometric meta-analysis for a cumulative science and practice of human resource management," *Human Resource Management Review*, vol. 27, no. 1, pp. 201–215, 2017.

[69] J. E. Hunter and F. L. Schmidt, *Methods of meta-analysis: Correcting error and bias in research findings*. Sage, 2004.

[70] M. W. Lipsey and D. B. Wilson, *Practical meta-analysis.* SAGE publications, Inc, 2001.

[71] M. Harrer, P. Cuijpers, T. Furukawa, and D. Ebert, *Doing meta-analysis with R: A hands-on guide*. Chapman and Hall/CRC, 2021.

[72] J. J. Deeks, J. P. Higgins, D. G. Altman, and C. S. M. Group, *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley Online Library, 2019, ch. Analysing data and undertaking meta-analyses, pp. 241–284.

[73] H. R. Rothstein, A. J. Sutton, and M. Borenstein, "Publication bias in meta-analysis," *Publication bias in meta-analysis: Prevention, assessment and adjustments*, pp. 1–7, 2005.

[74] W. Viechtbauer, "Conducting meta-analyses in R with the metafor, package," *Journal of Statistical Software*, vol. 36, no. 3, pp. 1–48, 2010.

[75] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International journal of social robotics*, vol. 1, pp. 71–81, 2009.

[76] A. Pereira, C. Martinho, I. Leite, and A. Paiva, "icat, the chess player: the influence of embodiment in the enjoyment of a game (short paper)," in *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems*, vol. 3, 2008, pp. 1253–1256.

[77] A. Powers, S. Kiesler, S. Fussell, and C. Torrey, "Comparing a computer agent with a humanoid robot," in *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, 2007, pp. 145–152.

[78] S. Kiesler, A. Powers, S. R. Fussell, and C. Torrey, "Anthropomorphic interactions with a robot and robot–like agent," *Social cognition*, vol. 26, no. 2, pp. 169–181, 2008.

[79] A. Krogsager, N. Segato, and M. Rehm, "Backchannel head nods in danish first meeting encounters with a humanoid robot: The role of physical embodiment," in *Human-Computer Interaction. Advanced Interaction Modalities and Techniques: 16th International Conference, HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part II 16*. Springer, 2014, pp. 651–662.

[80] K. Fischer, K. S. Lohan, and K. Foth, "Levels of embodiment: Linguistic analyses of factors influencing hri," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 2012, pp. 463–470.

[81] S. Thellman, M. De Graaf, and T. Ziemke, "Mental state attribution to robots: A systematic review of conceptions, methods, and findings," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 11, no. 4, pp. 1–51, 2022.