Crossing Reality: Comparing Physical and Virtual Robot Deixis

Zhao Han*
Yifei Zhu*
Colorado School of Mines
Golden, CO, USA
zhaohan@mines.edu,zhu1@mines.edu

Albert Phan
Fernando Sandoval Garza
Amia Castro
Colorado School of Mines
Golden, CO, USA

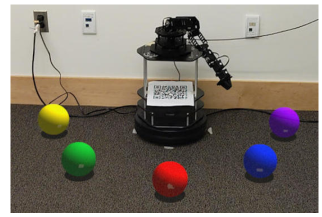
Tom Williams
Colorado School of Mines
Golden, CO, USA
twilliams@mines.edu

ABSTRACT

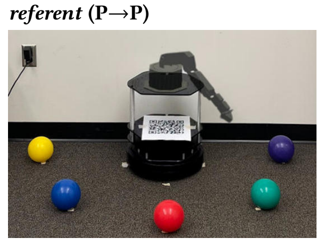
Augmented Reality (AR) technologies present an exciting new medium for human-robot interactions, enabling new opportunities for both implicit and explicit human-robot communication. For example, these technologies enable physically-limited robots to execute non-verbal interaction patterns such as deictic gestures despite lacking the physical morphology necessary to do so. However, a wealth of HRI research has demonstrated real benefits to physical embodiment (compared to, e.g., virtual robots on screens), suggesting AR augmentation of virtual robot parts could face challenges.

In this work, we present empirical evidence comparing the use of virtual (AR) and physical arms to perform deictic gestures that identify virtual or physical referents. Our subjective and objective results demonstrate the success of mixed reality deictic gestures in overcoming these potential limitations, and their successful use regardless of differences in physicality between gesture and referent. These results help to motivate the further deployment of mixed reality robotic systems and provide nuanced insight into the role of mixed-reality technologies in HRI contexts.

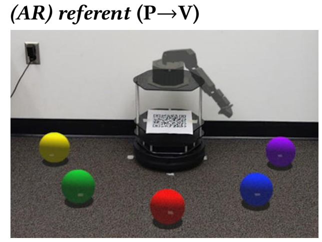
(a) Physical Robot with a *physical arm* pointing to a *physical referent* (P→P)



(b) Physical Robot with a *physical arm* pointing to an *virtual (AR) referent* (P→V)



(c) Physical Robot with a *virtual (AR) arm* pointing to a *physical referent* (V→P)



(d) Physical Robot with a *virtual (AR) arm* pointing to an *virtual (AR) referent* (V→V)

Figure 1: We investigate referring behavior at the intersection of physical and AR worlds (physical/virtual (AR) *arm* × physical/virtual (AR) *referent*).

CCS CONCEPTS

• **Computer systems organization → Robotics**; **External interfaces for robotics**; • **Human-centered computing → Mixed / augmented reality**; **Empirical studies in interaction design**.

KEYWORDS

Augmented reality (AR), deictic gesture, non-verbal communication, physical embodiment, presence, anthropomorphism, human-robot interaction (HRI)

ACM Reference Format:
Zhao Han, Yifei Zhu, Albert Phan, Fernando Sandoval Garza, Amia Castro, and Tom Williams. 2023. Crossing Reality: Comparing Physical and Virtual Robot Deixis . In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23), March 13–16, 2023, Stockholm, Sweden.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3568162.3576972

1 INTRODUCTION

In order to promote natural, human-like, and effective human-robot interactions, robots must be able to effectively communicate with people. Critically, this requires going beyond verbal communication alone. Due to robots' unique physical embodiment [88], human-robot interaction (HRI) researchers have investigated non-verbal behaviors [14], such as implicit arm movement (e.g., [23, 47]), gestures [77], and eye gaze [1, 59]. Multimodal approaches pairing these nonverbal displays with verbal communication have also been well-studied (e.g., [9, 37, 93]). These non-verbal behaviors, especially deictic gestures like pointing and presenting [77], are particularly important as they increase task efficiency [59] and improve subjective perceptions of robots [23].

Unfortunately, most robot systems (such as mobile or telepresence robots, autonomous vehicles, and free-flying drones) do not have the physical morphology to express many of these nonverbal cues, lacking heads and eyes for gazing, or arms for gesturing. Moreover, the high degree-of-freedom requirements and complex mechanics of these morphological components, especially physical arms, present cost barriers, especially when such components would only be used for gesturing and not for manipulation. Additionally, the inclusion of physical components like arms presents well-known safety concerns [31].

*Both authors contributed equally to this research.

HRI '23, March 13–16, 2023, Stockholm, Sweden

ACM ISBN 978-1-4503-9964-7/23/03.
https://doi.org/10.1145/3568162.3576972

To address these challenges, researchers have investigated *virtual* analogues to these traditional non-verbal cues. For nonverbal facial cues, this has taken a variety of forms. The Furhat robot head [2], for example, uses projection mapping to display a humanlike face without the need for precisely controlled animatronic facial parts. Similarly, many approaches use tablets to display robot faces (e.g., [28, 36]). Recently, AR technology has also been employed to visualize robot facial cues, allowing users or designers to customize expressions and easily change between facial expressions [101].

AR has also been recently used to provide a lower-cost solution for gestural capabilities. For example, Groechel et al. [31] studied the use of an AR arm on a mobile robot, and Hamilton et al. [32] and Brown et al. [11] compared AR arms to other types of AR annotations (e.g., arrows [93]). Results showed that arms were subjectively more well received. Yet the performance differences between virtual and physical arms have not yet been explored. As such, while the monetary cost differences between these options can be readily compared, the performance differences between these platforms are not yet well understood.

Moreover, it is unclear whether differences between virtual and physical arms might depend on the virtuality or physicality of the task-relevant objects to which a robot might choose to gesture. For example, virtual arms could be more effective (and viewed more positively) in tasks involving virtual referents, and vice versa, as a mismatch in physicality or virtuality of arm and referent could lead to confusion and delay. This would create a complex challenge given that mixed-reality task environments will necessarily contain a mixture of virtual and physical content, and the adoption of virtual appendages could be hindered by such concerns.

In this paper, we conducted a human-subjects study (N=36) to investigate the objective performance and subjective perception between physical and virtual (AR) arms, as mediated by the physicality or virtuality of the robot's target referent (See Figure 1). This work helps robot designers to better understand whether and when to employ virtual rather than physical morphological components. Moreover, this work provides insights that are sensitive to the nuances of mixed-reality robotics environments.

2 RELATED WORK

2.1 Physical Robots vs. Virtual Agents on Screen

Much HRI research has already demonstrated differences in objective performance and subjective perception between purely virtual and purely physical robotic entities, demonstrating that embodied physical presence leads to greater influence [3], learning outcomes [43], task performance [51, 89], gaze following from infants [56], proximity [3], exercise [25], positive perception [52], social facilitation [5], forgiveness [5], enjoyableness [25, 68, 88], helpfulness [25, 69], and social attractiveness [25]. However, these works compared entirely physical and entirely virtual robot presence, without considering morphologies that blend the physical and the virtual, as enabled by AR technologies.

2.2 Virtual Agents in AR

While in VR, virtual agents wholly reside within the virtual world, AR allows virtual objects and agents to be projected onto a user's view of the real world [91]. A variety of research has examined how interactants perceive agents in AR, and how space is perceived differently when the human user interacts with AR agents. Obaid et al. [63] showed that AR agents are perceived as physically distant by showing how participants adjust the volume of their speech, and Kim et al. [44] showed that AR agents that were aware of physical space were rated higher in terms of social presence.

Other researchers have examined how people perceive virtual humans (ostensibly) interacting with the physical world through AR. Lee et al. [48] studied AR-visualized humans subtly moving a physical table in terms of presence, co-presence, and attentional allocation, finding an increase in these measures. Schmidt et al. [79] experimented with virtual humans manipulating physical objects (e.g., hitting a physical ball with a virtual golf club), but did not find statistically significant differences in realism or emotional response. In contrast, our work considers a physical entity (a physical robot) with a virtual appendage, rather than a wholly virtual agent.

2.3 AR for Robot Communication

In this work, we are specifically interested in robots using AR appendages for communication. There has been a variety of work on the use of AR for human-robot communication within the broader area of VAM-HRI [90, 96]. Frank et al. [29] used AR to show reachable spatial regions in order to signal human users where and when to pass objects to robots. Taylor et al. [83] used AR to remove robot arm occlusion by making the arm transparent, and thus implicitly communicate the otherwise invisible context occluded by the arm. Diehl et al. [21] used AR to verify learned behavior in the robot learning domain to increase safety and trust. In addition to headset-based AR, researchers have also investigated projected AR. For example, Ganesan et al. [30] used projected AR to project car door frames and moving instructions in hopes of increasing task success in a car-assembly collaborative application. And Han et al use projector-based AR to communicate robotic intent [33–35].

2.4 Human and Robot Deictic Gesture

Finally, within this broader area, our work examines robots' use of AR visualizations for the purposes of deictic gestures. Deictic gesture has been a topic of sustained and intense study both in human-human interaction [49, 60] and human-robot interaction [77]. Deictic gesture is a key and natural communication modality, with humans starting to use deictic gesture around 9-12 months [7], and mastering it around age four [17]. Adults continue to use deixis to direct interlocutor attention, so as to establish joint and shared attention [54]. As a non-verbal modality, gesturing is especially helpful in public noisy environments such as factories, warehouses, or shopping malls when speech communication is not effective[14, 38].

Accordingly, roboticists have studied how deictic gestures can be applied to design more communicative robots, e.g., in tabletop environments [76] and free-form direction-giving scenarios [64]. Research shows that robots, like humans, can shift interlocutor attention [10] and can use a variety of deictic gestures, not limited to pointing [18, 77]. Williams et al. have begun to explore the use of deictic gesture within Augmented Reality [84, 93–95, 98], although most of this work has used non-anthropomorphic visualizations such as virtual arrows. In contrast, Hamilton et al. [32], like ourselves in this work, examine virtual arms, and show that AR virtual arms exhibit enhanced social presence and likability relative to virtual arrows – but also that the benefits of these approaches could be combined to gain the "best of both worlds" [11]. Unlike Hamilton et al., however, we are interested in explicitly comparing virtual arms to physical arms (rather than other types of virtual gestures) and we hope to better understand how the physical or virtual nature of the environment might mediate these differences.

2.5 Subjective Perceptions

Finally, we must discuss the specific dimensions of robot social perception that are of interest to us in this work.

2.5.1 Social Presence. Social presence (the feeling of being in the company of another social actor [81]) has been a central metric in studies involving virtual agents, as it can enable more effective social and group interactions [8, 53]. Within HRI, social presence has also been found to increase enjoyment and desire to re-interact [40]. It is unclear whether virtual robot appendages would make a robot seem less socially present than a physical robot appendage given the physicality of the robot's base. For similar reasons, it is unclear whether a robot's interactions with virtual objects would decrease a sense of social presence when the human themselves is also interacting with those virtual objects.

2.5.2 Anthropomorphism. Anthropomorphism is one of the most widely researched constructs within the HRI literature, and remains an area of extensive research [72, 73]. Projecting human characteristics to non-human entities [22, 24, 26, 100], such as attaching the AR virtual arm to the TurtleBot 2 in this work, encourages humans to re-use familiar interaction patterns from human-human interactions. This facilitates sensemaking and mental model alignment [22], leading humans to be more willing to interact, accept, and understand robot behaviors [46]. Robots that use gestures have been found to appear more anthropomorphic [75], and Hamilton et al. [32] specifically found that a mechanomorphic robot with a virtual arm may be viewed as more anthropomorphic (cf. [11]). However it is unclear whether virtual arms would be perceived as more or less anthropomorphic than physical arms.

2.5.3 Likability. As one of the primary metrics used in nonverbal robot communication [75, 94, 97], Likability summarizes peoples' overall perceptions of technology. Hamilton et al. [32] found evidence that virtual arms enhanced robot likability, but did not compare their approach with physical counterparts.

2.5.4 Warmth and Competence. As psychological constructs at the core of social judgment, warmth and competence are responsible for social perceptions among humans [27]. Warmth captures whether an actor is sociable and well-intentioned, and competence captures whether they can deliver on those intentions. Warmth and competence are thus key predictors of effective and preferable interactions, both for human-human interaction [27] and human-robot interaction [13, 78]. Moreover, they have been connected to social presence [39], and anthropomorphism [45, 92]. Because of these interrelations it is important to consider possible upstream or downstream effects on warmth and competence.

3 HYPOTHESES

Building on the body of related work described above, we thus formulate two key research questions:

- **RQ1:** Can a *virtual* robotic arm perform tasks as accurately and as efficiently as a *physical* robot arm while offering users a similarly natural interaction?
- **RQ2:** When there is a mismatch between the reality of the robot arm and the referent, will accuracy, efficiency, and subjective perception be affected?

Our work accordingly seeks to assess two sets of hypotheses.

3.1 Virtuality/Physicality of Robot Arms

First, we hypothesize that virtual robot arms in AR should not perform any worse than physical arms.

Hypothesis 1 (H1) – Virtual arms are just as accurate and efficient as Physical arms, or more so. We believe that robot deixis with a virtual arm will be no less accurate and efficient than deixis with a physical arm when identifying a referent. Efficiency will be measured by reaction time.

Hypothesis 2 (H2) – Virtual arms are perceived just as positively as Physical arms, or more so. Similarly, we believe that robot deixis with a virtual arm will be perceived equally or more positively (on dimensions like social presence, anthropomorphism, likability, warmth, and competence), than deixis with a physical arm when identifying a referent.

If these hypotheses hold, it will help to address potential concerns about the use of AR arms in mixed-reality environments, thus encouraging adoption of AR methods in future industrial contexts.

3.2 Reality Misalignment

Second, we hypothesize a mismatch in physicality/virtuality of Arm and the referent will have negative effects.

Hypothesis 3 (H3) – Reality misalignment negatively impacts users' objective ability to perform their tasks. Physical or virtual arms, when referring to physical or virtual referents, *respectively*, should have equivalent accuracy and efficiency. However, we hypothesize that a mismatch between these levels of reality (i.e., virtual arms pointing to physical objects, and vice versa) could decrease accuracy and efficiency, due to the need for additional cognitive processing to explicitly overcome this misalignment.

Hypothesis 4 (H4) – Reality misalignment negatively impacts users' subjective perceptions of their robotic teammates. Similarly, a mismatch between the reality of the arm and of the referent could negatively affect the user's subjective experience in identifying the robot arm's target.

If these hypotheses hold, it would provide guidance to robot developers and deployers of the types of contexts in which virtual gestures can and should be used.

4 METHOD

4.1 Apparatus

4.1.1 Robot Platform. Due to our interest in gesturally-limited robots, we used a TurtleBot 2 [65]. This differential wheeled robot is the second generation of the Turbot family, and is maintained by the current maintainer of Robot Operating System (ROS) [70], thus having a large support community. The specification for TurtleBot compatible platforms can be found on ros.org [99].

4.1.2 Augmented Reality Head-Mounted Display (AR-HMD). We use a Microsoft HoloLens 2 [57]: a commercial-grade see-through holographic mixed reality headset with a $43° \times 29°$ Field of View.

4.1.3 Physical/Virtual Robot Arm. The physical arm we used was a WidowX Robot Arm [71]: a 5-DoF arm with a parallel gripper, which can reach up to $41cm$ horizontally and $44cm$ vertically. Our virtual arm was created using the CAD models and Unified Robot Description Format (URDF) model of this arm [85], as rendered in Unity. The virtual arm has the same distance to the TurtleBot top when rendered in Unity. To affix the AR virtual arm to the same position as the physical arm, we placed a trackable QR code marker on the second top panel of the TurtleBot 2.

4.1.4 Physical/Virtual Referents. Five spheres were used as communicative referents and arced within the field of view of HoloLens 2 (See Figure 1). Each sphere measures $d = 15.24cm$ $(6in)$ in diameter and was placed $45°$ apart. The distance between the robot and each referents is $0.5m$ $(19.685in)$. The physical and rendered spheres had the same size and placement. To help perceive the location of virtual balls [20], shadows were added under them.

4.2 Gesturing Task and Implementation

The aforementioned materials were used in the context of a standard gesture-comprehension experiment. In each trial, the WidowX robot arm, mounted or simulated on top of a physical TurtleBot 2, randomly pointed to one of the colored spherical targets, which participants were then asked to identify by air-tapping on that target. This was repeated ten times, with targets chosen at random. While a controller could be used, the four directional buttons do not work well for five referents, and would introduce confounds for measuring response time.

For each gesture, the MoveIt motion planning framework [15] was used to move the robot's end effector to the desired pointing pose. As we conducted this experiment in person, the trajectory generated by MoveIt to move the robot arm to its final pose (which is traditionally non-deterministic due to the use of probabilistic algorithms [82]) was made deterministic by specifying a waypoint for which a deterministic trajectory could be guaranteed. This approach to a deterministic outcome has seen success in prior robot-to-human handover tasks and provides valuable experimental control in this new context [36].

For the AR virtual arm, we used ROS# [80] in Unity to receive the joint states to move the WidowX arm model rendered in Unity.

The MoveIt and Unity code is available on GitHub under the MIT licence to facilitate reproduction and replication: https://github.com/umhan35/ar-vs-physical-arm.

4.3 Experiment Design

This study followed a 2×2 *within-subjects* design. The ordering effect was counterbalanced using a full Latin square.

As implied throughout this work so far, we manipulated whether the arm and the referents, i.e., the spheres, are physical or rendered. Formally, two independent variables were manipulated: *Arm Physicality/Virtuality* and *Referent Physicality/Virtuality*. Thus, there were four study conditions across the two factors:

(1) P→P: *Physical* arm pointing at *Physical* spheres
(2) P→V: *Physical* arm pointing at *Virtual* spheres
(3) V→P: *Virtual* arm pointing at *Physical* spheres
(4) V→V: *Virtual* arm pointing at *Virtual* spheres

4.4 Procedure

After providing informed consent, participants completed a demographic survey and were randomly assigned to one of the four Latin Square orderings over the four experimental conditions. Participants watched three videos on how to wear HoloLens 2, run eye calibration, and use the air-tap gesture to confirm a target.

Then participants entered a sufficiently lighted experiment room to complete a practice round to get familiar with air-tapping. This practice allowed participants to walk through sample experiment trials to see how the robot arm moves to gesture, and practice air-tapping sphere targets. The practice round was also used to help mitigate novelty effects, as we expected that most participants would have no experience with a HoloLens 2 device. Experimenters asked clarifying questions to ensure participants' understanding of the task and the procedure during the practice round. During the experiment, participants stood $3m(9.84ft)$ away from the robot, so all spheres were in the field of view of the HoloLens 2.

Participants then began the experimental task. Each participant completed 10 trials in each condition. After completing each trial, participants were asked to answer a questionnaire containing our subjective measures. At the end of the experiment, participants were debriefed. It took 44.6 minutes on average to finish a study. This study design was approved by the human subjects research committee at Colorado School of Mines in USA.

4.5 Data Collection and Measures

To test our hypotheses, we collected two objective metrics and five subjective metrics, inspired in part by Hamilton et al. [32] and Brown et al. [11]. All experiment material, data, and analysis scripts are available at https://osf.io/27wbp/.

4.5.1 Objective Measures. Our objective metrics were collected using the air-tap and eye tracking [58] capabilities of the HoloLens 2. Participants were thus required to wear the HoloLens 2 in all conditions to ensure the same experiment settings, including when observing physical arms pointing to physical referents. Specifically, we used these two capabilities to collect two key objective measures. **Accuracy** was calculated as the percentage of true positives where participants "clicked" a target referent (by air-tap gesture). **Reaction time** was calculated as the duration between when the robot arm began moving from its home position to when participants looked at the target object. For conditions with physical balls, invisible balls were added in Unity at the location of the physical balls to use Unity's eye tracking capabilities.

4.5.2 Subjective Measures. The four key subjective measures discussed in Section 2 were collected using surveys administered after each experimental block. **Social Presence** was measured using the

Almere Social Presence scale [40]. **Anthropomorphism** was measured using the Godspeed Anthropomorphism scale [6]. **Likability** was measured using the Godspeed Likability scale [6]. **Warmth and Competence** were measured using the ROSAS Scale [13].

4.6 Data Analysis

We used the Bayesian analysis framework [87] to analyze our data, due to a number of benefits of Bayesian analysis over the more common Frequentist approach [87]. Most critical for us is the ability not simply to determine whether a null hypothesis can be rejected (as in the Frequentist approach), but rather to quantify evidence *for* and *against* competing hypotheses. That is, we are interested in the possibility of equivalence between certain conditions, and would want to collect evidence in favor of such an eventuality, and the Bayesian approach allows us to quantify evidence in favor of a lack of an effect ($\mathcal{H}_0$) just as easily as it allows us to quantify evidence in favor of the existence of an effect ($\mathcal{H}_1$), and provides (through Bayes Factor analysis) easily interpretable means of quantifying the relative strength of evidence (odds ratio) of one hypothesis relative to the other ($BF_{10} = 1/BF_{01}$). Note as well that the p value cannot provide a measure of evidence in favor of $\mathcal{H}_0$.

Our Bayesian approach also informed our recruitment strategy. While in the frequentist approach a power analysis is needed, in part because one is not permitted to "peek" at their data before sampling has concluded to decide whether to stop early or to extend sampling beyond initial intent [4, 12], this is not the case for Bayesian analysis because Bayesian analysis is not grounded in the central limit theorem. As such, it does not require power analysis [19], and experimenters can use flexible sampling plans in which data is collected until firm claims can be made or resources are exhausted. For more details, we refer readers to [87].

Within this analysis framework, we used version 0.16.3 of the JASP statistical software [42] to perform Bayesian Repeated-Measures Analyses of Variance with Random Slopes [86] and Bayes Factor Analysis [74], in which Bayes Inclusion Factors across matched models were computed using Bayesian Model Averaging [41, 55]. When a main effect or interaction effect could not be ruled out (i.e., the Inclusion Bayes Factor BF_{10} in favor of including the main or interaction term was above 0.333, or in other words, the Exclusion Bayes Factor BF_{01} against inclusion of the main or interaction term was below 3.0), post-hoc Bayesian t-tests were used to examine pairwise comparisons between conditions. In this paper, we always report Bayes Factors in the direction of our evidence. That is, when evidence favors exclusion of an effect, we report the Exclusion Bayes Factor BF_{01} (e.g., 3.5) rather than the equivalent Inclusion Bayes Factor BF_{10} (e.g., 0.286) for ease of readability.

4.7 Participants

45 participants were recruited at Colorado School of Mines in USA. Data from nine participants was excluded. One could not finish due to difficulty performing the air tap gesture. There was a networking problem for four participants, and the other four participants accidentally repeated conditions. Of the remaining 36 participants, 24 identified as male and 12 identified as female. Ages ranged from 18 to 40 (M=23.0, SD=5.19). 18 (50%) reported experience with robots, 6 (16.7%) were neutral, and 12 disagreed. 12 (33.3%) reported experience with augmented reality, 6 were neutral, and 18 (50%) disagreed. Each was given a US $15 Amazon gift card for participation.

5 RESULTS

We will now discuss the results for each of our measures. These results are summarized in Table 1.

5.1 Objective Measures

5.1.1 *Accuracy.*

A two-way repeated measures Analysis of Variance (RM-ANOVA) [86] was used to assess the effect of Arm and Referent Physicality/Virtuality on accuracy. This analysis revealed moderate evidence against an effect of Arm Physicality/Virtuality (BF_{01} = 4.011) (that is, our data was approximately 4 times more likely under models *excluding* an effect of Arm Physicality/Virtuality ($\mathcal{H}_0$) than under models *including* such an effect ($\mathcal{H}_1$)). This analysis also revealed moderate evidence against an effect of Referent Physicality/Virtuality (BF_{01} = 3.278). Finally, this analysis revealed anecdotal evidence against an interaction between Arm Physicality/Virtuality and Referent Physicality/Virtuality (BF_{01} = 1.261).

Because an interaction effect could not be ruled out, post-hoc Bayesian t-tests were used to perform pairwise comparisons between conditions. However, this analysis revealed anecdotal to moderate evidence against all pairwise differences ($BF_{01} \in [2.152, 4.795]$).

5.1.2 *Efficiency.*

An RM-ANOVA revealed moderate against an effect of Arm Physicality/Virtuality (BF_{01} = 4.519). This analysis also revealed moderate evidence against an effect of Referent Physicality/Virtuality (BF_{01} = 3.156). Finally, this analysis revealed anecdotal evidence against an interaction between Arm Physicality/Virtuality and Referent Physicality/Virtuality (BF_{01} = 1.650).

Because an interaction effect could not be ruled out, post-hoc Bayesian t-tests were used to perform pairwise comparisons between conditions. However, this analysis revealed anecdotal to moderate evidence against all pairwise differences ($BF_{01} \in [1.530, 5.470]$).

5.2 Subjective Measures

5.2.1 *Social presence.*

Before analyzing social presence, we conducted a Bayesian reliability analysis [66, 67] of our Almere social presence scale data. For McDonald's ω, the posterior mean equaled 0.794 with 95% CI=[0.738, 0.845]. For Cronbach's α, the posterior mean equaled 0.798 with 95% CI=[0.743, 0.847][1]. We thus calculated an unweighted composite score for each participant.

Mean social presence ratings were relatively low, with condition means ranging from 2.1 to 2.3. An RM-ANOVA revealed anecdotal against an effect of Arm Physicality/Virtuality (BF_{01} = 2.017), suggesting there probably is no such effect, but if there was, it would be that the Virtual Arm conveyed more social presence (M=2.228, SD=0.867) than the Physical Arm (M=2.122, SD=0.800). This analysis also revealed anecdotal evidence against an effect of Referent Physicality/Virtuality (BF_{01} = 2.845), suggesting again there probably is no such effect, but if there was, it would be that the Virtual Referent conveyed more social presence (M=2.208, SD=0.803) than the Physical Referent (M=2.142, SD=0.866). More data would need to be collected to fully rule out such effects. Finally, this analysis revealed

[1] Nunnally [62]'s widely-adopted recommended level is near 0.8.

Table 1: Means and Standard Deviations (SD) for all measures

	Arm Physicality/Virtuality		Referent Physicality/Virtuality		Arm/Referent Physicality/Virtuality			
Measure	*Physical Arm*	*Virtual Arm*	*Physical Referent*	*Virtual Referent*	*P⟶P*	*P⟶V*	*V⟶P*	*V⟶V*
Accuracy (%)	0.979±0.059	0.983±0.050	0.981±0.055	0.982±0.055	0.983±0.045	0.974±0.070	0.978±0.064	0.989±0.032
Reaction Time (s)	4.927±1.095	4.878±1.051	4.973±1.097	4.832±1.043	4.906±1.068	4.949±1.135	5.041±1.136	4.715±0.945
*Social Presence**	2.122±0.800	2.228±0.867	2.142±0.866	2.208±0.803	2.094±0.840	2.150±0.768	2.189±0.901	2.267±0.843
*Anthropormorph.**	2.214±0.847	2.681±0.835	2.467±0.877	2.428±0.859	2.294±0.909	2.133±0.784	2.639±0.842	2.722±0.838
*Likability**	3.414±0.657	3.575±0.757	3.469±0.694	3.519±0.731	3.417±0.651	3.411±0.673	3.522±0.740	3.628±0.780
*Warmth**	2.280±0.660	2.477±0.742	2.340±0.705	2.417±0.712	2.264±0.688	2.296±0.640	2.417±0.722	2.537±0.766
*Competence**	3.656±0.805	3.750±0.946	3.639±0.895	3.767±0.859	3.574±0.795	3.639±0.827	3.604±0.995	3.896±0.883

* Subjective measures were rated on a 5-item Likert scale.

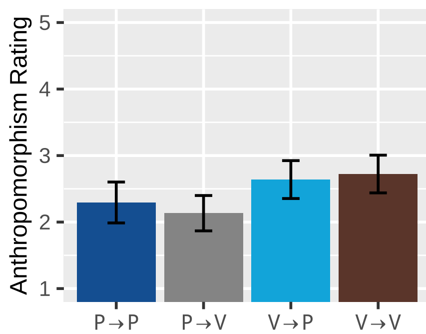


Figure 2: Mean anthropomorphism ratings. Error bars show 95% credible interval. Results shows Arm Virtuality/Physicality has an effect on anthropomorphism but not Referent Virtuality/Physicality.

moderate evidence against an interaction between Arm Physicality/Virtuality and Referent Physicality/Virtuality (BF_{01} = 4.257).

5.2.2 Anthropomorphism. Bayesian reliability analysis of the 5-item Godspeed [6] Anthropomorphism scale yielded ω = 0.799 (95% CI=[0.747, 0.851]), α = 0.802 (95% CI=[0.749, 0.855]).

As shown in Figure 2, mean anthropomorphism ratings were relatively low, with condition means ranging from 2.1 to 2.7. An RM-ANOVA revealed strong evidence for an effect of Arm Physicality/Virtuality (BF_{10} = 17.679), suggesting that the Virtual Arm was perceived as more anthropomorphic (M=2.681, SD=0.835) than was the Physical Arm (M=2.214, SD=0.847). This analysis also revealed moderate evidence against an effect of Referent Physicality/Virtuality (BF_{01} = 4.0816). Finally, this analysis revealed anecdotal evidence for an interaction between Arm Physicality/Virtuality and Referent Physicality/Virtuality (BF_{10} = 1.277).

Because an interaction effect could not be ruled out, post-hoc Bayesian t-tests were used for pairwise comparisons between conditions. These post-hoc t-tests revealed that while Virtual Arms were viewed as equally anthropomorphic (BF_{01} = 3.521) when gesturing towards Physical Referents (M=2.639, SD=0.832) and Virtual Referents (M=2.722, SD=0.838), Physical Arms may have been perceived as less anthropomorphic (BF_{10} = 1.532) when gesturing towards Physical Referents (M=2.294, SD=0.909) than when gesturing towards Virtual Referents (M=2.133, SD=0.784). That is, the combination of a Physical Arm gesturing towards a Physical Referent may have been uniquely non-anthropomorphic, but more data would be needed to confirm this difference.

5.2.3 Likability. Bayesian reliability analysis of the 5-item Godspeed [6] Likability scale yielded ω = 0.864 (95% CI=[0.831, 0.900]), α = 0.870 (95% CI=[0.838, 0.903]).

Mean likability ratings were relatively high, with condition means ranging from 3.4 to 3.6. An RM-ANOVA revealed anecdotal evidence against an effect of Arm Physicality/Virtuality (BF_{01} = 1.503), suggesting there probably is no such effect, but if there was, it would be that the Virtual Arm was viewed as more likable (M=3.575, SD=0.757) than the Physical Arm (M=3.414, SD=0.657). More data would be needed to rule out such an effect. This analysis also revealed moderate evidence against an effect of Referent Physicality/Virtuality (BF_{01} = 3.604). Finally, this analysis revealed anecdotal evidence against an interaction effect between Arm Physicality/Virtuality and Referent Physicality/Virtuality (BF_{01} = 2.280).

Because an interaction effect could not be ruled out, post-hoc Bayesian t-tests were used to perform pairwise comparisons between conditions. However, this analysis revealed anecdotal to moderate evidence against all pairwise differences ($BF_{01} \in [1.945, 5.574]$).

5.2.4 Warmth. Bayesian reliability analysis of the 6-item ROSAS [13] Warmth scale yielded ω = 0.807 (95% CI=[0.757, 0.850]), α = 0.799 (95% CI=[0.749, 0.847]).

Mean warmth ratings were relatively low, with condition means ranging from 2.4 to 2.6. An RM-ANOVA revealed anecdotal evidence against an effect of Arm Physicality/Virtuality (BF_{01} = 1.145), suggesting there probably is no such effect, but if there was, it would be that the Virtual Arm was viewed as more Warm (M=2.477, SD=0.742) than the Physical Arm (M=2.280, SD=0.660). More data would be needed to rule out such an effect. This analysis also revealed moderate evidence against an effect of Referent Physicality/Virtuality (BF_{01} = 3.105). Finally, this analysis also revealed moderate evidence against an interaction between Arm Physicality/Virtuality and Referent Physicality/Virtuality (BF_{01} = 3.328).

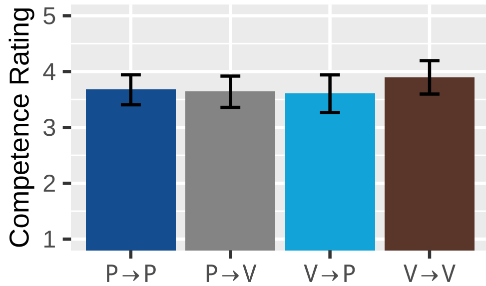


Figure 3: Mean competence ratings. Error bars show 95% credible interval. Arm had no effect on competence. For Referent Virtuality, there is no difference when a physical arm points (P→P/P→V). Additionally, there is no difference between different reality misalignments (P→V/V→P).

5.2.5 Competence. Bayesian reliability analysis of the 4-item ROSAS [13] Competence scale yielded ω = 0.824 (95% CI=[0.775, 0.867]), α = 0.827 (95% CI=[0.781, 0.870]).

As shown in Figure 3, mean competence ratings were relatively high, with condition means ranging from 3.6-3.9. An RM-ANOVA revealed moderate evidence against an effect of Arm Physicality/Virtuality (BF_{01} = 3.086). This also revealed anecdotal evidence against an effect of Referent Physicality/Virtuality (BF_{01} = 2.315), suggesting there probably is no such effect, but if there was, it would be that the robot was perceived as more competent when gesturing towards Virtual Referents (M=3.767, SD=0.859) than when gesturing towards Physical Referents (M=3.639, SD=0.895). Finally, this revealed anecdotal evidence for an interaction between Arm Physicality/Virtuality and Referent Physicality/Virtuality (BF_{10} = 1.596).

Because an interaction effect could not be ruled out, post-hoc Bayesian t-tests were used to perform pairwise comparisons between conditions. These post-hoc t-tests revealed that while Physical Arms were viewed as equally competent (BF_{01} = 5.423) when gesturing towards Physical Referents (M=3.574, SD=0.795) and Virtual Referents (M=3.639, SD=0.827), Virtual Arms may have been perceived as less competent (BF_{10} = 2.976) when gesturing towards Physical Referents (M=3.604, SD=0.995) than when gesturing towards Virtual Referents (M=3.896, SD=0.883). That is, the combination of a Virtual Arm gesturing towards a Virtual Referent may have been perceived as uniquely competent, but more data would be needed to confirm this difference.

6 DISCUSSION

6.1 Hypothesis One

Our first hypothesis was that Virtual arms would be just as accurate and efficient as Physical arms, or more so. Our results support this hypothesis. Overall, regardless of whether a physical or virtual arm was used, participants were highly accurate and equally efficient. This should provide assurance for robot designers concerned about accuracy and efficiency of augmented reality arms. That is, in contexts where robots' arms are only used for the purposes of deictic gesturing, it may be more cost effective for augmented reality visualizations to be used than physical arms, if a robot's deployment context already requires the use of a Mixed Reality headset.

One potential caveat, however, is the possibility that our results were due to ceiling effects. When designing this task, we were concerned that participants may have found it challenging due to lack of prior experience with Augmented Reality. And in fact, only 12 (33.3%) of our participants reported prior experience with AR. But despite this lack of prior experience, participants achieved over 97% accuracy in all conditions. It is possible that differences between conditions could have been more readily apparent in a more challenging task involving more objects, objects that are closer together, or objects for which the human participant would need to turn their head to follow the robot's deixis. For future work, we suggest examining AR and physical gestures in more collaborative tasks like identifying targets during assembly tasks.

6.2 Hypothesis Two

Our second hypothesis was that Virtual arms would be perceived just as positively as Physical arms, or more so. Our results support this hypothesis. Specifically, our results suggest that Virtual arms were perceived just as (or more) positively in terms of each of our five key metrics. That is, when the robot had a Virtual arm, participants viewed it as just as competent, as *or more* likable, socially present, warm, and distinctly more anthropomorphic.

We believe these effects are interrelated, and likely stem from the perceived differences in anthropomorphism. In this work, the robot in all conditions had relatively low levels of anthropomorphism (See Figure 2), but these levels were more moderate when the Virtual arm was used. There are several possible antecedents of this effect. First, the mechanical nature of the robot was less obvious when the virtual arm was used, both due to the lack of a physical mechanism in real-space and due to the lack of sounds from the robot's motors. In other words, the Physical robot may have suffered penalties to anthropomorphism due to the underlying truth to its mechanical nature. Second, it is possible that the animation of the Virtual arm appeared more fluid; or any motion disfluidities were less apparent.

Previous work [61] has shown that very low and very high levels of anthropomorphism are negatively correlated with social presence, but that moderate levels of anthropomorphism are positively correlated with social presence. This could explain why the Virtual Arm induced more social presence. Previous work [40, 72, 75] has similarly shown that Anthropomorphism and Social Presence lead to greater likability. It is also reasonable to expect that middling levels of Anthropomorphism may lead to greater perceived Warmth, just as they lead to greater perceived Social Presence. Finally, while there is evidence that Anthropomorphism plays a key role in mediating competence-based trust [16], it is unsurprising in our particular task that both robots were perceived as overall competent in their task, due to their shared abilities and limitations.

Synthesizing these trends, we believe (1) the non-mechanical nature of the Virtual arms circumvented certain penalties to anthropomorphism that would have otherwise occurred; (2) the resulting middling level of anthropomorphism for these Virtual Arms led to increased Social Presence and Warmth; (3) increased Social Presence and Warmth led to downstream effects on Likability.

Future work should confirm both our hypothesized explanations (e.g., relating to the auditory and visual components of the robot's behavior) and the hypothesized down-stream chain of effects stemming from differences in Anthropomorphism, some links of which are backed by specific work in the HRI community, some by work beyond the community, and some by hypothesis and intuition.

6.3 Hypothesis Three

Our third hypothesis was that reality misalignment (when a Physical Arm was used to gesture to a Virtual Referent, or a Virtual Arm was used to gesture to a Physical Referent) would negatively impact users' objective ability to perform their tasks. Our results did not support this hypothesis, with no interaction effects found between Arm Virtuality/Physicality and Referent Virtuality/Physicality for either of our objective measures. Despite our expectations, these results should thus provide assurance for robot designers considering using augmented reality visualizations to pick out physical objects, but also for those considering using physical arms to pick out virtual objects in mixed-reality tasks.

However, the same caveat applies here as for Hypothesis One. That is, since performance was uniformly good across the board, especially for Accuracy, it is possible our observations are due to ceiling effects, and that a more challenging task could have revealed differences. This represents an open direction for future work.

6.4 Hypothesis Four

Our fourth hypothesis was that reality misalignment would negatively impact users' subjective perceptions of their robotic teammates. While this hypothesis was not supported by our analysis, our analysis to assess this hypothesis revealed two intriguing effects.

6.4.1 Physicality Subverts Anthropomorphism. First, echoing the results found in service of H2, we observed that Physical Arms gesturing towards Physical objects were perceived as uniquely non-anthropomorphic. While one might think that the nature of the robot's environment would have little effect on perceptions of the robot itself, we believe that the concrete, grounded nature of Physical referents reinforced participants' perceptions of the Physical arm's physical embodiment and situated nature. As such, we interpret participants' perceptions of the robot's non-anthropomorphism in this case as perhaps not truly about a lack of anthropomorphism, so much as a gain in explicit mechanomorphism, or a sense of physical embodiment and groundedness.

If this is the case, it suggests new ways of measuring robot embodiment are needed, that de-emphasize a linear non-anthropomorphic to anthropomorphic spectrum, and instead emphasize either (a) placement within a multidimensional landscape of embodiment, or (b) the feeling of belonging-to-the-world, or of *perceived materiality*.

6.4.2 Virtuality Begets Competence within Virtuality. The second interesting finding arising from our analysis in service of this hypothesis was the observation that Virtual Arms *may* have been perceived as more competent in operating within the Virtual World, although the strength of our evidence falls just short of the threshold we would typically use to make such a claim with confidence. Future work could explore robotic performance of a variety of mixed-reality tasks that exhibit more range in terms of their analogy to real-world tasks. We suspect for example, that robots with virtual components may be perceived as especially capable and competent when performing tasks that are *inherently* virtual, such as creating, deleting, or adapting the properties of virtual objects, in ways that are simply not possible to perform for physical objects.

It is also possible that such changes in task context could also lead to differences in some of our other subjective metrics. A more social task context, for example, could have led to greater differences in perceived social presence. Moreover, this line of analysis raises interesting questions about the nature of social presence when two parallel worlds overlap and interact. Perhaps, for example, a robot may be perceived as having different degrees of social presence *with respect* to each of the physical and virtual worlds.

6.5 General Limitations

A final limitation that warrants discussion is our participant pool. Our participants were recruited from a uniquely pure-engineering university whose participants may have been predisposed to favor interaction modalities that highlighted mixed reality dimensions of interaction. Across higher education, but especially at Engineering schools, attendees not only have outsized experience with robotic and AR technologies, but moreover are systematically biased towards technological solutions to social problems [50]. As such, it is reasonable to suspect that our participant sample may have been predisposed to look favorably on technological configurations that seemed to leverage the widest swathe of the very technologies their educational context had conditioned them to value.

7 CONCLUSION

In this paper, we investigated the differences in performance between robots that gesture using either Physical or Virtual arms, as mediated by the physical or virtual nature of their gestural referents. Our results provide support for the utility of cost-saving AR technologies for human-robot communication, with no downsides observed to the use of Augmented Reality as a medium for nonverbal communication. Moreover, our results demonstrate that there is limited need for developers to worry about reality misalignment effects, especially when they are using Virtual arms. This further demonstrates the potential use of virtual robotic appendages even when interacting with the physical world. Finally, our results highlight key opportunities for future HRI research to pursue more nuanced study of both mixed-reality HRI, as well as foundational topics like Anthropomorphism whose importance extends beyond Mixed Reality domains but whose nuances are amplified therein.

ACKNOWLEDGMENTS

This work was supported in part by NSF grant IIS-1909864.

REFERENCES

[1] Henny Admoni and Brian Scassellati. 2017. Social eye gaze in human-robot interaction: a review. *Journal of Human-Robot Interaction* 6, 1 (2017), 25–63.

[2] Samer Al Moubayed, Jonas Beskow, Gabriel Skantze, and Björn Granström. 2012. Furhat: a back-projected human-like robot head for multiparty human-machine interaction. In *Cognitive behavioural systems*. Springer, 114–130.

[3] Wilma A Bainbridge, Justin W Hart, Elizabeth S Kim, and Brian Scassellati. 2011. The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics* 3, 1 (2011), 41–52.

[4] Madeleine E Bartlett, CER Edmunds, Tony Belpaeme, and Serge Thill. 2022. Have I Got the Power? Analysing and Reporting Statistical Power in HRI. *ACM Transactions on Human-Robot Interaction (THRI)* 11, 2 (2022), 1–16.

[5] Christoph Bartneck. 2003. Interacting with an embodied emotional character. In *Proceedings of the 2003 international conference on Designing pleasurable products and interfaces*. 55–60.

[6] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics* 1, 1 (2009), 71–81.

[7] Elizabeth Bates. 1976. *Language and context: The acquisition of pragmatics*. Academic Press.

[8] Frank Biocca, Chad Harms, and Judee K Burgoon. 2003. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & virtual environments* 12, 5 (2003), 456–480.

[9] Paul Bremner and Ute Leonards. 2016. Iconic gestures for robot avatars, recognition and integration with speech. *Frontiers in psychology* 7 (2016), 183.

[10] Andrew G Brooks and Cynthia Breazeal. 2006. Working with robots and objects: Revisiting deictic reference for achieving spatial common ground. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. 297–304.

[11] Landon Brown, Jared Hamilton, Zhao Han, Albert Phan, Thao Phung, Eric Hansen, Nhan Tran, and Tom Williams. 2022. Best of Both Worlds? Combining Different Forms of Mixed Reality Deictic Gestures. *ACM Transactions on Human-Robot Interaction* (2022).

[12] Katherine S Button, John Ioannidis, Claire Mokrysz, Brian A Nosek, Jonathan Flint, Emma SJ Robinson, and Marcus R Munafò. 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews neuroscience* 14, 5 (2013), 365–376.

[13] Colleen M Carpinella, Alisa B Wyman, Michael A Perez, and Steven J Stroessner. 2017. The Robotic Social Attributes Scale (RoSAS) Development and Validation. In *Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction*. 254–262.

[14] Elizabeth Cha, Yunkyung Kim, Terrence Fong, Maja J Mataric, et al. 2018. A survey of nonverbal signaling methods for non-humanoid robots. *Foundations and Trends® in Robotics* 6, 4 (2018), 211–323.

[15] Sachin Chitta, Ioan Sucan, and Steve Cousins. 2012. Moveit![ros topics]. *IEEE Robotics & Automation Magazine* 19, 1 (2012), 18–19.

[16] Lara Christoforakos, Alessio Gallucci, Tinatini Surmava-Große, Daniel Ullrich, and Sarah Diefenbach. 2021. Can robots earn our trust the same way humans do? A systematic exploration of competence, warmth, and anthropomorphism as determinants of trust development in HRI. *Frontiers in Robotics and AI* 8 (2021).

[17] Eve V Clark and CJ Sengul. 1978. Strategies in the acquisition of deixis. *Journal of child language* 5, 3 (1978), 457–475.

[18] Herbert H Clark. 2005. Coordinating with each other in a material world. *Discourse studies* 7, 4-5 (2005), 507–525.

[19] Joshua Correll, Christopher Mellinger, Gary H McClelland, and Charles M Judd. 2020. Avoid Cohen's 'small','medium', and 'large'for power analysis. *Trends in Cognitive Sciences* 24, 3 (2020), 200–207.

[20] Catherine Diaz, Michael Walker, Danielle Albers Szafir, and Daniel Szafir. 2017. Designing for depth perceptions in augmented reality. In *2017 IEEE international symposium on mixed and augmented reality (ISMAR)*. IEEE, 111–122.

[21] Maximilian Diehl, Alexander Plopski, Hirokazu Kato, and Karinne Ramirez-Amaro. 2020. Augmented reality interface to verify robot learning. In *Proceedings of the 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. 378–383.

[22] Carl DiSalvo and Francine Gemperle. 2003. From seduction to fulfillment: the use of anthropomorphic form in design. In *Proceedings of the 2003 international conference on Designing pleasurable products and interfaces*. 67–72.

[23] Anca D Dragan, Shira Bauman, Jodi Forlizzi, and Siddhartha S Srinivasa. 2015. Effects of robot motion on human-robot collaboration. In *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 51–58.

[24] Brian R Duffy. 2002. Anthropomorphism and robotics. *The society for the study of artificial intelligence and the simulation of behaviour* 20 (2002).

[25] Juan Fasola and Maja J Matarić. 2013. A socially assistive robot exercise coach for the elderly. *Journal of Human-Robot Interaction* 2, 2 (2013), 3–32.

[26] Julia Fink. 2012. Anthropomorphism and human likeness in the design of robots and human-robot interaction. In *International Conference on Social Robotics*. Springer, 199–208.

[27] Susan T Fiske, Amy JC Cuddy, and Peter Glick. 2007. Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences* 11, 2 (2007), 77–83.

[28] Naomi T Fitter and Katherine J Kuchenbecker. 2016. Designing and assessing expressive open-source faces for the Baxter robot. In *International Conference on Social Robotics*. Springer, 340–350.

[29] Jared A Frank, Matthew Moorhead, and Vikram Kapila. 2017. Mobile mixed-reality interfaces that enhance human–robot interaction in shared spaces. *Frontiers in Robotics and AI* 4 (2017), 20.

[30] Ramsundar Kalpagam Ganesan, Yash K Rathore, Heather M Ross, and Heni Ben Amor. 2018. Better teaming through visual cues: how projecting imagery in a workspace can improve human-robot collaboration. *IEEE Robotics & Automation Magazine* 25, 2 (2018), 59–71.

[31] Thomas Groechel, Zhonghao Shi, Roxanna Pakkar, and Maja J Matarić. 2019. Using socially expressive mixed reality arms for enhancing low-expressivity robots. In *Proceedings of the 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. 1–8.

[32] Jared Hamilton, Thao Phung, Nhan Tran, and Tom Williams. 2021. What's The Point? Tradeoffs Between Effectiveness and Social Perception When Using Mixed Reality to Enhance Gesturally Limited Robots. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. 177–186.

[33] Zhao Han, Jenna Parrillo, Alexander Wilkinson, Holly A Yanco, and Tom Williams. 2022. Projecting Robot Navigation Paths: Hardware and Software for Projected AR. In *2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI), Short Contributions*.

[34] Zhao Han, Alexander Wilkinson, Jenna Parrillo, Jordan Allspaw, and Holly A Yanco. 2020. Projection mapping implementation: Enabling direct externalization of perception results and action intent to improve robot explainability. In *2020 AAAI Fall Symposium on The Artificial Intelligence for Human-Robot Interaction (AI-HRI)*.

[35] Zhao Han, Tom Williams, and Holly A Yanco. 2022. Mixed-Reality Robot Behavior Replay: A System Implementation. In *2022 AAAI Fall Symposium on The Artificial Intelligence for Human-Robot Interaction (AI-HRI)*.

[36] Zhao Han and Holly Yanco. 2019. The effects of proactive release behaviors during human-robot handovers. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 440–448.

[37] Zhao Han and Holly A Yanco. 2022. Communicating Missing Causal Information to Explain a Robot's Past Behavior. *ACM Transactions on Human-Robot Interaction (THRI)* (2022).

[38] Simon Harrison. 2011. The creation and implementation of a gesture code for factory communication. In *GESPIN 2011: Gesture and speech in interaction*.

[39] Khaled Hassanein and Milena Head. 2007. Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping. *International Journal of Human-Computer Studies* 65, 8 (2007), 689–708.

[40] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. 2010. Assessing acceptance of assistive social agent technology by older adults: the almere model. *International journal of social robotics* 2, 4 (2010), 361–375.

[41] Max Hinne, Quentin F Gronau, Don van den Bergh, and Eric-Jan Wagenmakers. 2020. A conceptual introduction to Bayesian model averaging. *Advances in Methods and Practices in Psychological Science* 3, 2 (2020), 200–215.

[42] JASP Team. 2022. JASP (Version 0.16.3)[Computer software]. https://jasp-stats.org/

[43] James Kennedy, Paul Baxter, and Tony Belpaeme. 2014. Children comply with a robot's indirect requests. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. 198–199.

[44] Kangsoo Kim, Gerd Bruder, and Greg Welch. 2017. Exploring the effects of observed physicality conflicts on real-virtual human interaction in augmented reality. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. 1–7.

[45] Seo Young Kim, Bernd H Schmitt, and Nadia M Thalmann. 2019. Eliza in the uncanny valley: anthropomorphizing consumer robots increases their perceived warmth but decreases liking. *Marketing letters* 30, 1 (2019), 1–12.

[46] Dieta Kuchenbrandt, Nina Riether, and Friederike Eyssel. 2014. Does anthropomorphism reduce stress in HRI?. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. 218–219.

[47] Minae Kwon, Sandy H Huang, and Anca D Dragan. 2018. Expressing robot incapability. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 87–95.

[48] Myungho Lee, Kangsoo Kim, Salam Daher, Andrew Raij, Ryan Schubert, Jeremy Bailenson, and Greg Welch. 2016. The wobbly table: Increased social presence via subtle incidental movement of a real-virtual table. In *2016 IEEE Virtual Reality (VR)*. IEEE, 11–17.

[49] Stephen C Levinson. 2004. 5 Deixis. *The handbook of pragmatics* (2004), 97.

[50] Jon A Leydens and Juan C Lucena. 2017. *Engineering justice: Transforming engineering education and practice*. John Wiley & Sons.

[51] Daniel Leyzberg, Samuel Spaulding, Mariya Toneva, and Brian Scassellati. 2012. The physical presence of a robot tutor increases cognitive learning gains. In *Proceedings of the annual meeting of the cognitive science society*, Vol. 34.

[52] Jamy Li. 2015. 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* 77 (2015), 23–37.

[53] Matthew Lombard and Theresa Ditton. 1997. At the heart of it all: The concept of presence. *Journal of computer-mediated communication* 3, 2 (1997).

[54] Max M Louwerse and Adrian Bangerter. 2005. Focusing attention with deictic gestures and linguistic expressions. In *27th Annual Conference of the Cognitive Science Society*. 1331–1336.

[55] S Mathôt. 2017. Bayes like a Baws: Interpreting Bayesian repeated measures in JASP. *Cognitive Science and more. Retrieved from: https://www. cogsci. nl/blog/interpreting-bayesian-repeated-measures-in-jasp* (2017).

[56] Andrew N Meltzoff, Rechele Brooks, Aaron P Shon, and Rajesh PN Rao. 2010. "Social" robots are psychological agents for infants: A test of gaze following. *Neural networks* 23, 8-9 (2010), 966–972.

[57] Microsoft. 2019. *About HoloLens 2.* Retrieved 2022-02-15 from https://docs.microsoft.com/en-us/hololens/hololens2-hardware

[58] Microsoft. 2019. *Eye tracking on HoloLens 2.* Retrieved 2022-02-15 from https://docs.microsoft.com/en-us/windows/mixed-reality/design/eye-tracking

[59] AJung Moon, Daniel M Troniak, Brian Gleeson, Matthew KXJ Pan, Minhua Zheng, Benjamin A Blumer, Karon MacLean, and Elizabeth A Croft. 2014. Meet me where i'm gazing: how shared attention gaze affects human-robot handover timing. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction.* 334–341.

[60] Sigrid Norris. 2011. Three hierarchical positions of deictic gesture in relation to spoken language: a multimodal interaction analysis. *Visual Communication* 10, 2 (2011), 129–147.

[61] Kristine L Nowak and Frank Biocca. 2003. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments* 12, 5 (2003), 481–494.

[62] Jum C. Nunnally. 1994. *Psychometric theory.* Tata McGraw-Hill Education.

[63] Mohammad Obaid, Radosław Niewiadomski, and Catherine Pelachaud. 2011. Perception of spatial relations and of coexistence with virtual agents. In *International Workshop on Intelligent Virtual Agents.* Springer, 363–369.

[64] Yusuke Okuno, Takayuki Kanda, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. 2009. Providing route directions: design of robot's utterance, gesture, and timing. In *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI).* 53–60.

[65] Open Source Robotics Foundation. 2022. *TurtleBot2: Open-source robot development kit for apps on wheels.* https://www.turtlebot.com/turtlebot2/

[66] Julius M Pfadt, Don van den Bergh, Klaas Sijtsma, and Eric-Jan Wagenmakers. 2022. A tutorial on Bayesian single-test reliability analysis with JASP. *Behavior Research Methods* (2022), 1–10.

[67] Julius M Pfadt, Don van den Bergh, Klaas Sijtsma, Morten Moshagen, and Eric-Jan Wagenmakers. 2021. Bayesian estimation of single-test reliability coefficients. *Multivariate Behavioral Research* (2021), 1–30.

[68] Kathrin Pollmann, Christopher Ruff, Kevin Vetter, and Gottfried Zimmermann. 2020. Robot vs. voice assistant: Is playing with pepper more fun than playing with alexa?. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction.* 395–397.

[69] Aaron Powers, Sara , Susan Fussell, and Cristen Torrey. 2007. Comparing a computer agent with a humanoid robot. In *Proceedings of the 2007 ACM/IEEE international conference on Human-robot interaction.* 145–152.

[70] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, Andrew Y Ng, et al. 2009. ROS: an open-source Robot Operating System. In *ICRA workshop on open source software*, Vol. 3. Kobe, Japan, 5.

[71] Trossen Robotics. [n. d.]. *WidowX Robot Arm Kit.* Retrieved 2022-02-12 from https://www.trossenrobotics.com/widowxrobotarm/

[72] Eileen Roesler, Dietrich Manzey, and Linda Onnasch. 2021. A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction. *Science Robotics* 6, 58 (2021), eabj5425.

[73] Eileen Roesler, Dietrich Manzey, and Linda Onnasch. 2022. Embodiment Matters in Social HRI Research: Effectiveness of Anthropomorphism on Subjective and Objective Outcomes. *ACM Transactions on Human-Robot Interaction* (2022).

[74] Jeffrey N Rouder, Richard D Morey, Paul L Speckman, and Jordan M Province. 2012. Default Bayes factors for ANOVA designs. *Journal of mathematical psychology* 56, 5 (2012), 356–374.

[75] Maha Salem, Friederike Eyssel, Katharina Rohlfing, Stefan Kopp, and Frank Joublin. 2013. To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics* 5, 3 (2013), 313–323.

[76] Maha Salem, Stefan Kopp, Ipke Wachsmuth, Katharina Rohlfing, and Frank Joublin. 2012. Generation and evaluation of communicative robot gesture. *International Journal of Social Robotics* 4, 2 (2012), 201–217.

[77] Allison Sauppé and Bilge Mutlu. 2014. Robot deictics: How gesture and context shape referential communication. In *Proceedings of the 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI).* 342–349.

[78] Marcus M Scheunemann, Raymond H Cuijpers, and Christoph Salge. 2020. Warmth and competence to predict human preference of robot behavior in physical human-robot interaction. In *Proceedings of the 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN).* 1340–1347.

[79] Susanne Schmidt, Oscar Javier Ariza Nunez, and Frank Steinicke. 2019. Blended agents: Manipulation of physical objects within mixed reality environments and beyond. In *Symposium on Spatial User Interaction.* 1–10.

[80] Siemens. 2022. *ROS# GitHub Repo.* Retrieved 2022-05-05 from https://github.com/siemens/ros-sharp/

[81] Paul Skalski and Ron Tamborini. 2007. The role of social presence in interactive agent-based persuasion. *Media psychology* 10, 3 (2007), 385–413.

[82] Ioan A Sucan, Mark Moll, and Lydia E Kavraki. 2012. The open motion planning library. *IEEE Robotics & Automation Magazine* 19, 4 (2012), 72–82.

[83] Ada V Taylor, Ayaka Matsumoto, Elizabeth J Carter, Alexander Plopski, and Henny Admoni. 2020. Diminished Reality for Close Quarters Robotic Telemanipulation. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* 11531–11538.

[84] Nhan Tran, Trevor Grant, Thao Phung, Leanne Hirshfield, Christopher Wickens, and Tom Williams. 2021. Get this?⇓ mixed reality improves robot communication regardless of mental workload. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction.* 412–416.

[85] Trossen Robotics. 2022. *The widowx arm description ROS package.* Retrieved 2022-02-15 from https://github.com/AnasIbrahim/widowx_arm/tree/master/widowx_arm_description

[86] Don van den Bergh, Eric-Jan Wagenmakers, and Frederik Aust. 2022. Bayesian Repeated-Measures ANOVA: An Updated Methodology Implemented in JASP. (2022).

[87] Eric-Jan Wagenmakers, Maarten Marsman, Tahira Jamil, Alexander Ly, Josine Verhagen, Jonathon Love, Ravi Selker, Quentin F Gronau, Martin Šmíra, Sacha Epskamp, et al. 2018. Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic bulletin & review* 25, 1 (2018), 35–57.

[88] Joshua Wainer, David J Feil-Seifer, Dylan A Shell, and Maja J Mataric. 2006. The role of physical embodiment in human-robot interaction. In *The 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).* IEEE, 117–122.

[89] Joshua Wainer, David J Feil-Seifer, Dylan A Shell, and Maja J Mataric. 2007. Embodiment and human-robot interaction: A task-based perspective. In *The 16th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).* IEEE, 872–877.

[90] Michael Walker, Thao Phung, Tathagata Chakraborti, Tom Williams, and Daniel Szafir. 2022. Virtual, Augmented, and Mixed Reality for Human-Robot Interaction: A Survey and Virtual Design Element Taxonomy. *arXiv preprint arXiv:2202.11249* (2022).

[91] Isaac Wang, Jesse Smith, and Jaime Ruiz. 2019. Exploring virtual agents for augmented reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* 1–12.

[92] Adam Waytz, Joy Heafner, and Nicholas Epley. 2014. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology* 52 (2014), 113–117.

[93] Tom Williams, Matthew Bussing, Sebastian Cabrol, Elizabeth Boyle, and Nhan Tran. 2019. Mixed reality deictic gesture for multi-modal robot communication. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI).* 191–201.

[94] Tom Williams, Matthew Bussing, Sebastian Cabrol, Ian Lau, Elizabeth Boyle, and Nhan Tran. 2019. Investigating the Potential Effectiveness of Allocentric Mixed Reality Deictic Gesture. In *Proceedings of the 11th International Conference on Virtual, Augmented, and Mixed Reality.*

[95] Tom Williams, Leanne Hirshfield, Nhan Tran, Trevor Grant, and Nicholas Woodward. 2020. Using augmented reality to better study human-robot interaction. In *International Conference on Human-Computer Interaction.* Springer, 643–654.

[96] Tom Williams, Daniel Szafir, Tathagata Chakraborti, and Heni Ben Amor. 2018. Virtual, augmented, and mixed reality for human-robot interaction. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction.* 403–404.

[97] Tom Williams, Daria Thames, Julia Novakoff, and Matthias Scheutz. 2018. "Thank You for Sharing that Interesting Fact!": Effects of Capability and Context on Indirect Speech Act Use in Task-Based Human-Robot Dialogue. In *Proceedings of the 13th ACM/IEEE International Conference on Human-Robot Interaction.*

[98] Tom Williams, Nhan Tran, Josh Rands, and Neil T. Dantam. 2018. Augmented, Mixed, and Virtual Reality Enabling of Robot Deixis. In *Proceedings of the 10th International Conference on Virtual, Augmented, and Mixed Reality.*

[99] Melonee Wise and Tully Foote. [n. d.]. *Specification for TurtleBot Compatible Platforms.* Retrieved 2022-02-12 from https://www.ros.org/reps/rep-0119.html

[100] Jakub Złotowski, Diane Proudfoot, Kumar Yogeeswaran, and Christoph Bartneck. 2015. Anthropomorphism: opportunities and challenges in human–robot interaction. *International journal of social robotics* 7, 3 (2015), 347–360.

[101] Sebastian Meyer zu Borgsen, Patrick Renner, Florian Lier, Thies Pfeiffer, and Sven Wachsmuth. 2018. Improving human-robot handover research by mixed reality techniques. *VAM-HRI* (2018), 2018–03.