Best of Both Worlds? Combining Different Forms of Mixed Reality Deictic Gestures

LANDON BROWN, JARED HAMILTON, ZHAO HAN, ALBERT PHAN, THAO PHUNG, ERIC HANSEN, NHAN TRAN, and TOM WILLIAMS, Colorado School of Mines, USA

Mixed Reality provides a powerful medium for transparent and effective human-robot communication, especially for robots with significant physical limitations (e.g., those without arms). To enhance nonverbal capabilities for armless robots, this article presents two studies that explore two different categories of mixed reality deictic gestures for armless robots: a virtual arrow positioned over a target referent (a non-ego-sensitive allocentric gesture) and a virtual arm positioned over the gesturing robot (an ego-sensitive allocentric gesture). In Study 1, we explore the tradeoffs between these two types of gestures with respect to both objective performance and subjective social perceptions. Our results show fundamentally different task-oriented versus social benefits, with non-ego-sensitive allocentric gestures enabling faster reaction time and higher accuracy, but ego-sensitive gestures enabling higher perceived social presence, anthropomorphism, and likability. In Study 2, we refine our design recommendations by showing that in fact these different gestures should not be viewed as mutually exclusive alternatives, and that by using them together, robots can achieve both task-oriented and social benefits.

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

Additional Key Words and Phrases: Augmented Reality (AR), Mixed Reality (MR), deictic gesture, nonverbal communication, social presence, anthropomorphism, Human-Robot Interaction (HRI)

ACM Reference format:
Landon Brown, Jared Hamilton, Zhao Han, Albert Phan, Thao Phung, Eric Hansen, Nhan Tran, and Tom Williams. 2023. Best of Both Worlds? Combining Different Forms of Mixed Reality Deictic Gestures. *ACM Trans. Hum.-Robot Interact.* 12, 1, Article 9 (February 2023), 23 pages.
https://doi.org/10.1145/3563387

1 INTRODUCTION

For robots to communicate effectively with humans, they must be capable of natural, human-like human-robot dialogue [9, 44, 58]. And, in contrast to dialogue agents and chatbots, interactive robots must be able to communicate with sensitivity to situated context [44, 63]. This requires three broad competencies: *environmental context sensitivity* (sensitivity to the spatially situated,

In addition, Nhan Tran is now affiliated with Cornell University, Ithaca, NY, 14853.

L. Brown, J. Hamilton, Z. Han, A. Phan, and T. Phung contributed equally to this work.

This work was funded in part by National Science Foundation grants IIS-1909864 and CNS-1810631.

Authors' address: L. Brown, J. Hamilton, Z. Han (corresponding author), A. Phan, T. Phung, E. Hansen, N. Tran, and T. Williams (corresponding author), MIRRORLab, Department of Computer Science, Colorado School of Mines, 1500 Illinois St., Golden, CO, 80401; emails: lcbrown@mines.edu, jaredghamilton1998@gmail.com, zhaohan@mines.edu, aphan@mines.edu, phungpthao@gmail.com, ejhansen@mines.edu, nt322@cornell.edu, twilliams@mines.edu.

2573-9522/2023/02-ART9 $15.00
https://doi.org/10.1145/3563387

large-scale, uncertain, and incompletely known nature of task environments [75]); *cognitive context sensitivity* (sensitivity to the working memory and attentional constraints of teammates [76]); and *social context sensitivity* (sensitivity to the relational context into which they are embedded, and the importance of strengthening and maintaining social relationships through adherence to social and moral norms [41, 79] and building of trust and rapport [18, 33, 52]).

For these three competencies to be mastered, robots must be able to understand and generate both verbal behaviors and nonverbal behaviors such as gesture and eye gaze. Nonverbal behaviors are critical for situated interaction [2, 17, 29, 49] and are integrally related to these three competencies. Deictic gestures such as pointing leverage environmental context by identifying nearby referents, especially when such referents are not currently known or attended to by interlocutors. These gestures are often generated due to cognitive context to direct interlocutor attention [42] and reduce memory costs that would be otherwise imposed by communication [15, 49]. And gestures are often generated with sensitivity to social context by mimicking the gestures of interlocutors to increase engagement and build rapport through mirroring [8].

While there has been a host of research on nonverbal behavior generation in **Human-Robot Interaction (HRI)** [1, 6, 7, 51, 54–56], the ability for robots to use these human-like attention-directing nonverbal cues has been traditionally dependent on a robot's morphology, with gaze typically requiring eyes or heads, and gesture typically requiring arms. This is potentially problematic as many robots (both social and nonsocial) lack these morphological elements. Moreover, for many robots, the addition of these elements not only may be difficult to justify on the basis of affordability but also may be simply infeasible (e.g., it may be infeasible to add a gesture-capable arm to a UAV due to payload constraints). Accordingly, researchers have been investigating new methods for nonverbal signaling (e.g., directed lighting cues) that may achieve those goals typically addressed by physical gaze and gesture [11, 61] without requiring the same types of hardware-intensive morphological constraints. One family of methods that has attracted substantial recent attention consists of *Mixed Reality Deictic Gestures*; gestures or gesture-like cues visualized using **Mixed Reality Head-Mounted Displays (MR-HMDs)** such as the Microsoft HoloLens and the Magic Leap.

Mixed Reality Deictic Gestures are a form of view-augmenting Mixed Reality Interaction Design Elements, as viewed through the lens of the Reality-Virtuality Interaction Cube framework by Williams et al. [77]. Mixed Reality Deictic Gestures are classified into at least four primary types: allocentric, perspective-free, ego-sensitive allocentric, and ego-sensitive perspective-free gestures [80]. Allocentric gestures point out the communicator's target referent using imagery seen only from the viewer's perspective (e.g., circling a target referent within a user's Mixed Reality head-mounted display). Perspective-free gestures are generated onto the environment from a third-party perspective (e.g., projecting a circle around a target referent on the floor of the shared environment). Ego-sensitive allocentric gestures are generated only from the viewer's perspective, except they connect the communicator to its referent (e.g., pointing to a target referent using a simulated arm rendered in a user's MR-HMD), and ego-sensitive perspective-free gestures are gestures generated by a third party onto the environment that also connect the communicator to its referent (e.g., projecting a line from the robot to its target on the floor of the shared environment). Within the Virtual Design Element Taxonomy presented by [68], all of these forms of visualizations can be categorized as *Robot Comprehension Visualizations* → *Entity* → *Entity Locations*. This is, for example, how some work in this space [72] has been categorized in the public visualization of this taxonomy (https://vam-hri.mybluemix.net/). However, some forms of ego-sensitive allocentric gestures (e.g., virtual arms) may also be regarded as *Virtual Alterations* → *Morphological* → *Body Extensions*. In previous work, we specifically investigated the first of these categories, allocentric gestures, and demonstrated that mixed reality gestures can significantly increase the communicative effectiveness of nonhumanoid robots [65, 72, 73].

One downside of these previous explorations of allocentric gestures is the low ecological validity of the evaluation context, with crowdworkers viewing interactive videos simulating the expected appearance of such gestures. Accordingly, participants in previous experiments had full Field of View and viewed the entire experimental environment through an unchanging vantage point. In realistic task contexts, users are unlikely to be able to view their entire task environment from a single perspective, and Mixed Reality deictic gestures must be delivered through platforms like the HoloLens or Magic Leap, limiting the portion of the environment in which gestures can be displayed. In even moderately larger task contexts, these factors could result in users completely directing their Field of View and attention toward the regions where mixed reality deictic gestures are being displayed, completely avoiding the nonhumanoid robot generating the visualizations. This lack of attention toward the robot could have detrimental long-term effects on human-robot teaming, such as decreased trust, rapport, and situation awareness.

These challenges may be addressable by another form of Mixed Reality deictic gesture highlighted in the Williams et al. [80] taxonomy: ego-sensitive allocentric gestures, in which simulated arms are rendered above the robot and used to point just as physical arms would, as seen in the work of Groechel et al. [30]. The use of such arms could increase the robot's anthropomorphism, and because users would need to consistently look toward the robot to see where it is pointing, such arms could increase the robot's social presence.

On the other hand, ego-sensitive allocentric gestures may come with their own challenges. Because users will need to follow the vector along which the robot is pointing and estimate which objects fall within the robot's deictic cone, they may be less accurate and efficient at determining the targets of those gestures, especially when target referents are far from the robot (the very context in which ego-sensitive allocentric gestures are expected to provide social benefits).

In **Study 1**[1] (Section 3), we systematically investigated these expected differences and found tradeoffs between ego-sensitive allocentric gestures (an anthropomorphic arm positioned above a robot) and non-ego-sensitive allocentric gestures (a virtual arrow positioned above target referents), specifically in terms of tradeoffs between these forms with respect to objective and subjective measures. We observed that ego-sensitive allocentric gestures (virtual arms) led to increases in subjective, socially oriented measures, such as the anthropomorphism and social presence of the robot, while non-ego-sensitive allocentric gestures (virtual arms) led to increases in objective, task-oriented measures, such as accuracy and reaction time.

As we will discuss, we believed that these differences in Study 1 may have been due in part to differences in gaze requirements, as virtual arms required interlocutors to consistently look toward the robot to see where it was pointing, increasing social presence while impeding objective task-oriented performance, especially for referents far from the robot (for which following the robot's deictic cone would present greater difficulty for interlocutors). Meanwhile, when virtual arrows were used, interlocutors could more or less ignore the robot, as the gestures appeared directly over target referents, maximizing task-oriented performance while undermining the robot's subjective perception.

Certain elements of the design of Study 1 may have limited the ability to assess this particular hypothesis, however. In the experiment, there was limited difference between cases in which referents were considered to be close to the robot (1 meter) versus far from the robot (2 meters). In addition, we used the Hololens 1, whose limited Field of View may have artificially imposed the sorts of visual behaviors they described. As such, we argue that greater confidence could be

[1] This first study first appeared in a paper presented at HRI 2021 [32].

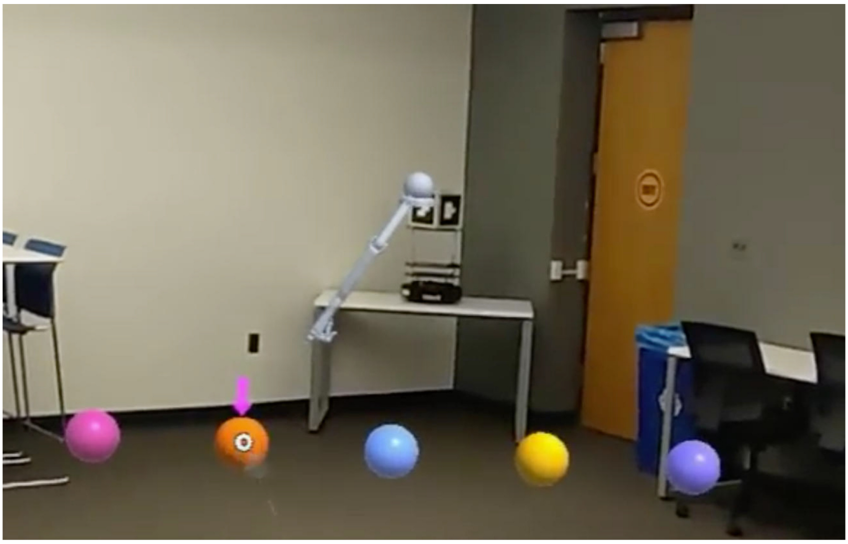


Fig. 1. A condition from Study 2 in which a robot uses both ego-sensitive (virtual arm) and non-ego-sensitive (virtual arrow) allocentric gestures to point toward gestural targets close to itself, whereas Study 1 shows these two gestures separately. The two studies show that, by using those gestures together, robots can achieve both task-oriented and social benefits.

gathered in the differences found in Study 1 if an HMD with a wider field of view (the HoloLens 2) were used, and if there were more obvious separation between distance-based conditions.

Moreover, while virtual arms and virtual arrows were present in Study 1 as competing options for nonverbal robotic communication, it is unclear whether these two types of gestures need to be viewed as mutually exclusive options. We argue that it may well be the case that these gesture types can be used together in a way that obtains the benefits of both approaches while ameliorating the limitations of either approach. We envision that when used together, a robot may be viewed more positively socially due to the use of the virtual arm, and because virtual arrows would clearly pick out the objects being pointed to by that virtual arm, it would be able to achieve those subjective social benefits without sacrificing objective task performance. Coincidentally, Goktan et al. [28] have also explored some of these concerns after we reported Study 1 while part of our work below was under review.

In **Study 2** (Section 4), we thus evaluated these expectations through a within-subjects laboratory experiment and used the HoloLens 2 with a wider Field of View. Our results suggested that by combining ego-sensitive allocentric gestures (virtual arms) and non-ego-sensitive allocentric gestures (virtual arrows) (as shown in Figure 1), we are able to simultaneously achieve some—but not all—of the subjective and objective benefits previously attributed to these two types of gestures.

2 RELATED WORK

2.1 Human and Robot Deictic Gesture

Deixis is a key component of human-human communication [39, 46]. Humans begin pointing while speaking even from infancy, using deictic gestures around 9 to 12 months [4], and mastering deictic reference around age 4 [13]. Among adults, deictic gestures remain a critical component of situated communication, helping direct interlocutor attention in order to establish joint and shared attention [2]. Deictic gestures also help humans express their thoughts, especially in environments in which verbal communication would be difficult, such as in noisy factory environments [34]. Accordingly, HRI researchers have been investigating how to enable effective robotic understanding and generation of deictic gestures.

Widespread evidence has been found in the HRI literature for the effectiveness of robots' use of nonverbal cues, including deictic gestures such as *pointing*, across a variety of different contexts, including tabletop environments [55] and free-form direction-giving [48]. Not only can robots use deictic gestures just as effectively as humans for the purpose of shifting interlocutor attention [7], but also it has been shown that robots' use of deictic gestures also improves subsequent human recall and human-robot rapport [6]. Sauppé and Mutlu [56], for example, investigated a group of robotic deictic gestures: touching, presenting, grouping, pointing, sweeping, and exhibiting. This group of gestures was inspired by Clark [14], who studied human deictic gestures and concluded that humans use more than just pointing as deictic gestures. Just as Sauppé and Mutlu [56] examined the objective and subjective differences between these six categories of *physical* deictic gestures, researchers have recently begun to examine the objective and subjective differences between different categories of *Mixed Reality* deictic gestures.

2.2 Mixed Reality Human-robot Communication

While the utility of robot deictic gestures has been well demonstrated, the generation of these gestures by robots is subject to a number of constraints. First, the ability to generate precise and natural deictic gestures typically requires robot arms: a morphological choice that not only is often prohibitively expensive but also does not make sense for all use cases due to factors such as arm size and weight (e.g., in the case of drones). In order to enable gestural capabilities while avoiding the financial and morphological limitations of traditional deictic gesture, some researchers have recently been investigating the use of *Mixed Reality* gestures that could serve these same purposes. Before describing these new forms of gestures, we will briefly describe the scope of recent work being conducted on Mixed Reality for Human-Robot Interaction (see also [31, 68, 77, 78]).

Mixed Reality provides opportunities for a host of new forms of human-robot communication across domains such as task-driven interaction [26, 27, 59], collaboration exploration [50], and social interaction [12, 69]. For example, Frank et al. [26] present an MR interface that shows a robot's reachable spatial regions in green and unreachable regions in red so that humans can better understand where and how to pass objects to the robot. In [27], humans and robots work together to assemble and move a car door. In the MR interface, the robot gives the human teammate information about its working area, what part of the door it will work on next, moving instructions, and the success of each subtask. In [12], the robot projects its trajectory and the spatial region that it would occupy while moving on the floor so the nearby humans can see and make a move to avoid collision with the robot.

Even for referential communication alone, Augmented and Mixed Reality afford many new forms of communication. For example, Sibirtseva et al. [59] present a Mixed Reality interface in which circles and labels are used to indicate the set of objects a robot is considering during reference resolution. By looking at those candidate objects, humans can provide additional explanation to the robot to assist disambiguation. In contrast, Reardon et al. [50] present an MR interface for collaborative human-robot exploration tasks. Once a robot finds a target object, a trajectory from the human's current location to that object's location is visualized in the MR interface, to aid navigation to that object.

While Sibirtseva et al. [59] focus on *passive backchannel communication* and Reardon et al. [50] focus on *passive nonlinguistic communication*, we are particularly interested in *active linguistic communication* in which Mixed Reality imagery can be visualized alongside spoken language as an alternative to physical gesture. In previous work [72], we presented a framework for categorizing the

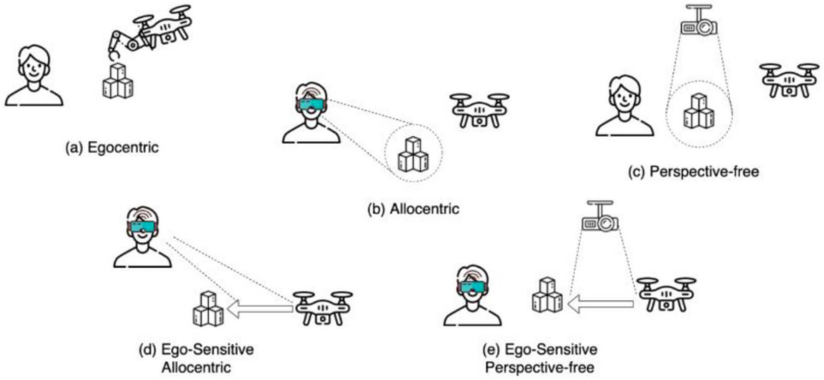

(a) Egocentric
(b) Allocentric
(c) Perspective-free
(d) Ego-Sensitive Allocentric
(e) Ego-Sensitive Perspective-free

Fig. 2. Five categories of mixed reality gestures [72] (used with permission).

different types of Mixed Reality deictic gestures that could be used in this way, delineating between four main categories of deictic gestural cues unique to Mixed Reality environments (Figure 2):

(1) **Perspective-free gestures:** gestures that are projected onto the environment from a third-party perspective. Weng et al. [71], for example, studied where a robot should project arrows onto a tabletop to reference target objects.
(2) **Ego-sensitive perspective-free gestures:** gestures that are projected onto an environment in a way that connects the communicator (i.e., the robot) to its referent, e.g., if Weng et al. [71] had generated their visualizations in such a way that the base of generated arrows originated at the robot.
(3) **Allocentric gestures:** gestures that pick out the communicator's target referent using imagery generated from the viewer's perspective (e.g., within MR-HMDs). Previously, we [72, 73] prototyped one category of Mixed Reality deictic gesture, non-ego-sensitive allocentric gestures (e.g., gestures like circles and arrows, generated from a user's perspective, without taking the robot generator into account), and provided the first evidence for the effectiveness of those gestures within an online evaluation testbed. More recently, Tran et al. [64] demonstrated the effectiveness of these for the first time on realistic robotic and mixed reality hardware.
(4) **Ego-sensitive allocentric gestures:** gestures that connect the communicator to its referent within the viewer's perspective. For example, Groechel et al. [30] augmented an otherwise armless robot with virtual arms shown in an MR interface.

However, while researchers are beginning to prototype and explore gestures within each of these categories, there has been no previous research comparing gestures between these categories. In this work, we present the first such research, systematically comparing between the non-ego-sensitive and ego-sensitive categories of allocentric Mixed Reality deictic gestures. In Study 1, we investigated these two categories in particular because we believed they present a challenging tradeoff between objective task performance and subjective robotic perception, as detailed in the next section. In Study 2, we further investigated them because we believed these different gestures should not be viewed as mutually exclusive alternatives but should be used together.

2.3 Metrics of Success for Robot Deixis

While most research on robot deictic gestures has evaluated gestures based on *objective*, task-driven metrics such as accuracy and reaction time of gesture interpretation, recent work has also sought to evaluate how those gestures are *subjectively* perceived. Sauppé and Mutlu [56], for example, evaluate gestures on the basis of how *natural* they appear to be, and Williams et al. [72, 74] evaluate gestures on the basis of their impact on robot *likability*.

We believed that the two categories of gestures we examine may present a challenging design case in which each of the two gestural options is differentially beneficial with respect to a fundamentally different category of metric. First, we would expect non-ego-sensitive allocentric gestures (e.g., circles and arrows) to perform better on the objective measures delineated above. When a robot uses ego-sensitive allocentric gesures (e.g., virtual arms attached to its body), viewers must follow the deictic cone extending from the robot along its arm and attempt to determine which objects might fall within that cone. In contrast, when a robot uses non-ego-sensitive deictic gestures (e.g., circles and arrows), the robot's intended target is immediately and obviously picked out in the user's Field of View, providing little opportunity for inaccuracy or inefficiency. However, we expect that they may in turn perform better on *subjective* measures, such as anthropomorphism, social presence, likability, warmth, and competence.

Below, we will examine each of these categories and articulate why we believe the use of ego-sensitive allocentric gestures may lead to higher subjective ratings in those categories.

Anthropomorphism is the projection of human characteristics to nonhuman entities [20, 21, 23]. Within HRI, researchers have framed anthropomorphism in terms of the contrast between robots designed in the image of humans (with humanlike features and appendages) and robots designed in the image of animals (i.e., zoomorphism) or robots with purely functional designs [25]. Anthropomorphism has been shown to be valuable for robot design as it cues models of human-human interaction, facilitating sensemaking and mental model alignment [20]; leading humans to be more willing to interact with, accept, and understand robot's behaviors; and reducing human stress during interaction [38]. Moreover, robots that use gestures have been found to appear more anthropomorphic [54].

We expected robots using ego-sensitive allocentric gestures to be viewed as more anthropomorphic because they can provide human-like morphological features to otherwise mechanomorphic robots, can provide the illusion of motion and life to otherwise inert robots, and are more directly analogous to traditional physical robot gestures.

Social Presence is the feeling of being in the company of another social actor [60] and has been long explored within media studies due to the potential for a technology's social presence to enable more effective social and group interactions [5, 40]. Within HRI, researchers have found that robot social presence facilitates user enjoyment and desire to re-interact [36], perhaps due to our innate drive to seek out, engage in, and respond to socially interactive behaviors with other social actors. Social presence is also related to anthropomorphism in interesting ways. Specifically, Nowak and Biocca [47] found that very low and very high levels of anthropomorphism led to lower levels of social presence than middling levels of anthropomorphism.

We predicted that robots using ego-sensitive allocentric gestures will be viewed as more anthropomorphic, but not *highly* anthropomorphic, both due to their status as obvious augmentations and because the arms we explore in this work are not highly photorealistic renderings of human arms. We further predicted that this will lead robots to be perceived as having greater social presence. Moreover, we believed that the use of robot arms is likely to engender greater social presence due to the impact these arms may have on visual attention. That is, while cues like circles and arrows may be interpreted without looking at or considering the robot generating them, virtual arms require the user to repeatedly regard the robot in order to interpret its gestures. We believed this is likely to significantly increase the perceived presence of the robot.

Likability is a key usability metric used to summarize people's overall perceptions of technology, and has been one of the primary metrics used across the HRI field [3]. In gesture-related contexts, including in Mixed Reality contexts, it has been found that gesture use can lead to increased likability [54, 73, 79].

Because we predicted that robots using ego-sensitive allocentric gestures will be viewed as more anthropomorphic and as having higher social presence, warmth, and competence, we thus believed that this will then also lead them to be perceived as more likeable.

Finally, **Warmth and Competence** are social psychological constructs that are at the core of social judgment and are nearly entirely responsible for social perceptions among humans [24]. While warmth captures whether an actor is sociable and well intentioned, the competence captures whether they have the potential to deliver on those intentions. Warmth and competence are thus valuable within human-human interaction as they lead to more positive emotions [24]. Within HRI, warmth and competence have been found to be key predictors for human preferences between robots and robot behaviors [57] and have been shown to lead to more positive human-robot interactions [10]. These concepts are also related to those discussed above: warmth in particular is often associated with Social Presence [35], and anthropomorphism has been shown in certain contexts to directly lead to greater warmth [37] and competence-based trust [70].

Because we predicted that robots using ego-sensitive allocentric gestures will be viewed as more anthropomorphic and as having higher social presence, we thus believed that this will then also lead them to be perceived as more warm and competent.

3 STUDY 1

3.1 Hypotheses

Based on our review of the previous work discussed above, we formulate the following hypotheses:

H1: A robot that uses non-ego-sensitive allocentric gestures (i.e., arrows drawn over target referents) when referring to target referents will (**H1.1**) be *more effective* than a robot using ego-sensitive allocentric gestures (i.e., pointing using virtual arms) as measured by (1) accuracy and (2) reaction time, and (**H1.2:**) these benefits will be more pronounced for objects farther away from the robot.

H2: A robot that uses non-ego-sensitive allocentric gestures (i.e., arrows drawn over target referents) when referring to target referents will (**H2.1:**) have *lower social perception* than a robot using ego-sensitive allocentric gestures (i.e., pointing using virtual arms) as measured by (1) social presence, (2) anthropomorphism, (3) likability, (4) warmth, and (5) perceived competence, and (**H2.2:**) these detriments would be more pronounced for objects farther away from the robot.

3.2 Experiment

To investigate these hypotheses, we conducted a within-subjects human-subject experiment in which participants interacted with a robot in a mixed reality HRI context.

3.2.1 Experimental Design. Our experiment consisted of a series of four Latin-Square order-counterbalanced experiment blocks in each of which participants interacted with a robot in a mixed reality HRI context. In each of these experiment blocks, participants performed a *gesture understanding* task consisting of 10 trials. No time limit was imposed, but participants were encouraged to complete the task as quickly as possible. In each trial, participants' robot teammate gestured to one of three target referents (multi-colored spheres) at random using a Mixed Reality deictic gesture, of the form shown in Figures 3 and 4, depending on experimental condition, and participants were required to "click" (using an air-tap gesture) on the referent they believed the robot was gesturing toward. Between experiment blocks, participants completed surveys assessing their perceptions of the robot and its gestures, as described in Section 3.2.4.

Experimental blocks differed according to a 2 × 2 design in which two independent variables were manipulated: *Gesture Type* and *Referent Distance*. Specifically, each task was conducted in one

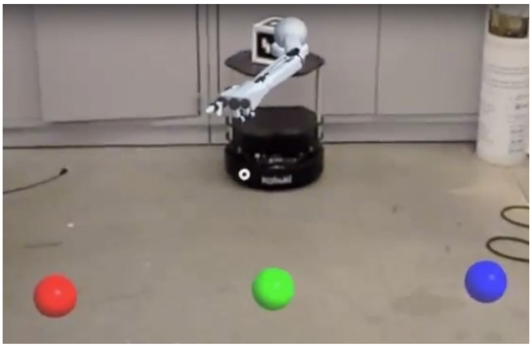


Fig. 3. Robot arm gesturing to holographic sphere (not in experimental environment).

of two *Gesture Type* conditions: in two of the four within-subject blocks, participants interacted with a robot that gestured using *ego-sensitive allocentric gestures* in which a virtual arm reached out and pointed toward target referents; in the other two within-subject blocks, participants interacted with a robot that gestured using *non-ego-sensitive allocentric gestures* in which an arrow appeared over target referents. Each of these two conditions was then further subdivided into two *Referent Distance* conditions: a *robot-close* condition in which the robot's target referents were approximately 1 meter from the robot and 2 meters from the human, and a *robot-distant* condition, in which the robot's target referents were approximately 2 meters from the robot and 1 meter from the human.

3.2.2 Experimental Apparatus. We designed and implemented a system that generated virtual entities such as the virtual robot arm and virtual arrow. The system includes three main physical components: the HoloLens, TurtleBot, and an MR cube.

Robotic Platform. A Kobuki TurtleBot 2 was used, affixed with an MR cube: a 12cm cardboard cube with fiducial markers on each face. This cube served as an anchor for the robot arm in the *arm* conditions and allowed the HoloLens to determine the robot's position in all conditions.

Mixed Reality Head-mounted Display. The MR-HMD used in this experiment was a Microsoft HoloLens, a commercial-grade stereographic Mixed-Reality Headset with a 30° × 17.5° Field of View. Participants' air-tap gestures were detected using the HoloLens's built-in gesture recognition capabilities.

Mixed Reality Deictic Gestures. Two mixed reality deictic gestures were designed in Unity: an *arrow* and *arm*. The arrow was a simple magenta arrow that statically appeared over target referents, shown in Figure 4. The arm used a virtual arm model created and textured using Blender and animated in Unity using a custom-built key-frame-based animation library.

Experimental Application. The experimental procedure and autonomous robot behavior were coded as a Unity application deployed onto the HoloLens. When viewing the scene through the HoloLens, participants perceived three spheres (red, green, and blue) hovering a half-meter above the ground between the subject and the TurtleBot. Different colors were used for compatibility with intended future work. The HoloLens also enabled participants to see the robot's gestures. In the *arm* conditions, an arm was always visible over the TurtleBot. In the *arrow* conditions, the arm was invisible and an arrow instead appeared over target referents.

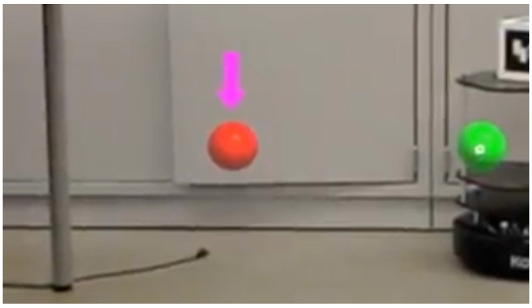


Fig. 4. Virtual arrow pointing to holographic sphere (the robot arm doesn't show up in this case).

3.2.3 Procedure. Upon arriving at the lab, participants provided informed consent and completed a demographic survey. Participants were then introduced to the TurtleBot, the HoloLens, and the task through both verbal instruction and an interactive tutorial designed in Unity and deployed on the HoloLens. The tutorial interface showed instruction text and virtual red, blue, and green spheres, walking participants through a sample experimental trial. During the tutorial, the participants learned how to use air-tap gestures to choose a sphere. After demonstrating the ability to successfully air-tap a target sphere three times, participants proceeded to the experiment. After completing the experiment, participants were paid $10 and debriefed.

3.2.4 Measures. To assess our two hypotheses, seven key metrics were collected during our experiment, including two objective measures and five subjective measures.

Objective Measures. Our first hypothesis was assessed using two objective measures:

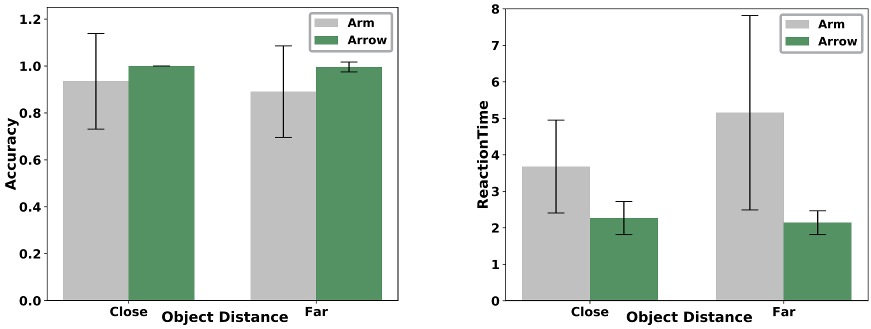
- **Accuracy** was measured as the percent of trials in which the target selected after a gesture was in fact the target of that gesture.
- **Reaction Time** was measured as the time (in seconds) from the time a gesture was triggered to the time a user selected the object they believed to be indicated by that gesture.

Subjective Measures. Our second hypothesis was assessed using five sets of survey questions administered after each experiment block. Each set of survey questions was a Likert scale composed of five to six items asking for statement agreement or disagreement on a 1–5 scale.

- **Social Presence** was measured using the Almere Social Presence scale [36].
- **Anthropomorphism** was measured using the Godspeed II Anthropomorphism scale [3].
- **Likability** was measured using the Godspeed II likability scale [3].
- **Warmth** was measured using the RoSAS Warmth scale [10].
- **Competence** was measured using the RoSAS Competence scale [10].

3.2.5 Participants. Twenty-four participants were recruited from the Colorado School of Mines through web postings and flyers (14 male, 10 female) for an ethics-board-approved experiment. Participants ranged in age from 18 to 52 (M = 22.46, SD = 7.86). Twenty of the 24 participants had not previously engaged in any experiments from our laboratory involving Mixed Reality. One participant failed to complete the experiment, leaving 23 usable data points.

3.2.6 Analysis. Data analysis was performed within a Bayesian analysis framework using the JASP 0.8.5.1 [62] software package, using the default settings justified by Wagenmakers et al. [67]. For each measure, a Bayesian **Repeated Measures Analysis of Variance (RM-ANOVA)** [16, 45, 53] was performed, using Gesture Type and Referent Distance as random factors. Bayes Inclusion



(a) Accuracy (95% CI)

(b) Reaction Time (95% CI)

Fig. 5. Objective results.

Factors Across Matched Models ("Baws Factors") [43] were then computed for each candidate main effect and interaction, indicating (in the form of a Bayes Factor) the evidence weight of all candidate models including that effect compared to the evidence weight of all candidate models not including that effect. Analysis of Likert scale data was performed after averaging responses within each scale.

3.3 Results

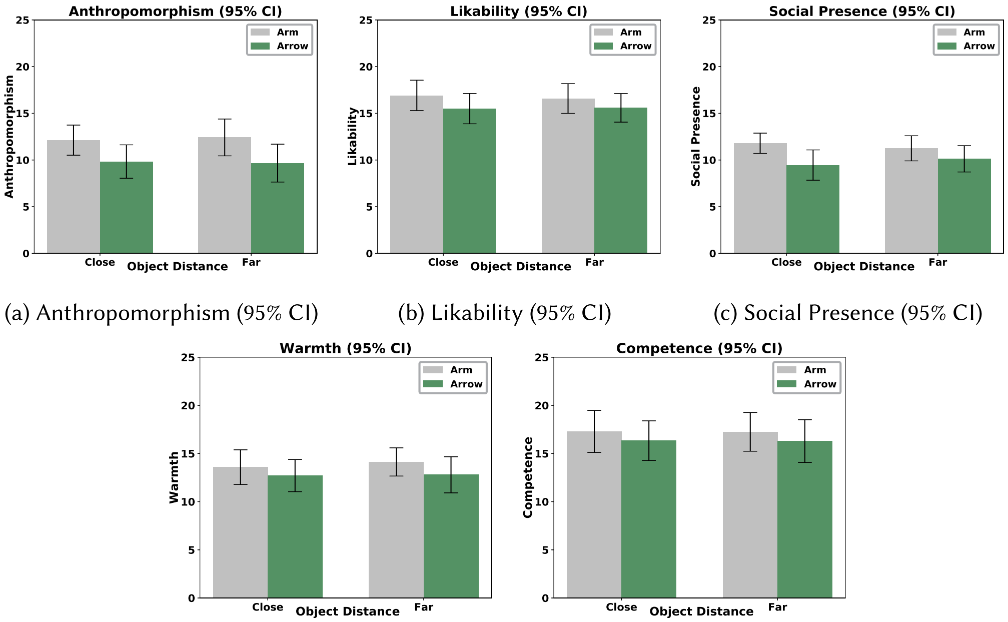
3.3.1 Objective Results. Figure 5 summarizes our main objective results.

Accuracy. Our results provided strong evidence in favor of an effect of Gesture Type on accuracy (**Bayes Factor (BF)** 16.376), as shown in Figure 5(a), suggesting that our data were 16 times more likely to be generated under models in which Gesture Type was included than under those in which it was not, and specifically that when virtual arrows were used, participants were more accurate in determining the intended target of those gestures (close distance (M = 1, SD = 0), far distance (M = 0.996, SD = 0.021)) than when virtual arms were used (close distance (M = 0.935, SD = 0.204), far distance (M = 0.891, SD = 0.195)). However, anecdotal evidence was found *against* an interaction effect between Gesture Type and Referent Distance on accuracy (BF 0.415), with the data 1/0.415 = 2.41 times less likely to have been generated under models including such an interaction.

Reaction time. Our results provided strong evidence in favor of an effect of Gesture Type on reaction time (BF 22.264), as shown in Figure 5(b), suggesting specifically that when virtual arrows were used, participants could more quickly identify the targets of those gestures (close distance (M = 2.265, SD = 1.047), far distance (M = 2.139, SD = 0.757) than when virtual arms were used (close distance (M = 3.678, SD = 2.941), far distance (M = 5.152, SD = 6.154)). However, anecdotal evidence was found *against* an interaction effect between Gesture Type and Referent Distance on reaction time (BF 0.455); that is, the data was 1/0.455 = 2.198 times less likely to have been generated under models including such an interaction.

3.3.2 Subjective Results. Figure 6 summarizes our main subjective results.

Social Presence. Our results provided extreme evidence in favor of an effect of Gesture Type on social presence (BF 440.332), as shown in Figure 6(c), suggesting specifically that when virtual arrows were used, participants viewed the robot as having lower social presence (close distance (M = 9.458, SD = 3.845), far distance (M = 10.125, SD = 3.327)) than when virtual arms were used (close distance (M = 11.792, SD = 2.570), far distance (M = 11.250, SD = 3.179)). However, our results



(a) Anthropomorphism (95% CI) (b) Likability (95% CI) (c) Social Presence (95% CI)

(d) Warmth (95% CI) (e) Competence (95% CI)

Fig. 6. Subjective results.

provided no significant evidence for or against an interaction between Gesture Type and Referent Distance on social presence, suggesting that more data must be collected[2] before a conclusion can be reached, which was unfortunately not allowed due to the COVID pandemic. As shown in Figure 6(c), it is plausible but not yet verifiable that when objects were *close* to the robot, use of virtual arms led to greater robotic social presence.

Anthropomorphism. Our results provided strong evidence in favor of an effect of Gesture Type on anthropomorphism (BF 6026.6), as shown in Figure 6(a), suggesting specifically that when virtual arrows were used, participants viewed the robot as having lower anthropomorphism (close distance (M = 9.833, SD = 4.239), far distance (M = 9.667, SD = 4.239)) than when virtual arms were used (close distance (M = 12.125, SD = 3.826), far distance (M = 12.417, SD = 4.671)). However, moderate evidence was found *against* an interaction effect between Gesture Type and Referent Distance on perceived anthropomorphism (BF 0.301); that is, the data was 1/0.301 = 3.322 times less likely to have been generated under models including such an interaction.

Likability. Our results provided moderate evidence in favor of an effect of Gesture Type on likability (BF 6.145), as shown in Figure 6(b), suggesting specifically that when virtual arrows were used, participants viewed the robot as having lower likability (close distance (M = 15.500, SD = 3.845), far distance (M = 15.583, SD = 3.263)) than when virtual arms were used (close distance (M = 16.917, SD = 3.855), far distance (M = 16.583, SD = 3.764)). However, moderate evidence was found *against* an interaction effect between Gesture Type and Referent Distance on perceived

[2] Within a Bayesian analysis framework, it is not only permitted but also actively encouraged to continue collecting data until a clear conclusion can be reached. This stands in stark opposition to Frequentist analyses, which depend on power analyses and rigid sampling plans.

likability (BF 0.319); that is, the data was 1/0.319 = 3.13 times less likely to have been generated under models including such an interaction.

Warmth. Our results provided no significant evidence for or against an effect of Gesture Type on warmth (BF 1.567), as shown in Figure 6(d), suggesting that more data must be collected before a conclusion can be reached. Moreover, moderate evidence was found *against* an interaction effect between Gesture Type and Referent Distance on perceived warmth (BF 0.328); that is, the data was 1/0.328 = 3.049 times less likely to have been generated under models including such an interaction.

Competence. Our results provided no significant evidence for or against an effect of Gesture Type on competence (BF 1.194), as shown in Figure 6(e), suggesting that more data must be collected before a conclusion can be reached. Moreover, moderate evidence was found *against* an interaction effect between Gesture Type and Referent Distance on perceived competence (BF 0.284); that is, the data was 1/0.284 = 3.521 times less likely to have been generated under models including such an interaction.

3.4 Discussion

3.4.1 Hypothesis One. We hypothesized that a robot that uses virtual arrows when referring to target referents would **(H1.1)** be more effective than a robot using virtual arms, as measured by accuracy and reaction time, and **(H1.2)** that these benefits would be more pronounced for objects farther away from the robot.

Our results support Hypothesis H1.1 but not Hypothesis H1.2. Our result suggests that a robot using virtual arrows is more effective than a robot using virtual arms: virtual arrows allowed users to complete the task faster and more accurately than virtual arms. This is unsurprising as virtual arrows directly pick out target referents without users needing to follow and interpret a deictic cone. While in the arrow scenario referent distance did not appear to impact accuracy and reaction time, in the arm scenario such an effect was observed: when the target referent was close to the robot, users could more accurately and quickly identify it.

While our Hypothesis H1.1 is supported, the results are inconclusive for Hypothesis H1.2. The anecdotal evidence against an interaction effect between Gesture Type and Referent Distance on accuracy (BF 0.415) and reaction time (BF 0.455) is not strong enough to conclusively rule out an effect, and visual inspection suggests there may indeed have been effects of distance on both accuracy and reaction time, in which task performance improved for virtual arms when referents were closer to the robot. More data will be needed to confirm or rule out these effects in a larger environment allowing greater distinction between distance conditions.

3.4.2 Hypothesis Two. We hypothesized that a robot that uses virtual arrows when referring to target referents would **(H2.1)** have lower social perception than a robot using arms as measured by social presence, anthropomorphism, likability, warmth, and perceived competence, and **(H2.2)** that these detriments would be more pronounced for objects farther away from the robot. We will thus separately assess this hypothesis for each of these subjective measures.

Our results support Hypothesis H2.1 but fail to support Hypothesis H2.2. First, our results suggest that robots using arms have higher social perception in terms of anthropomorphism, social presence, and likability than non-ego-sensitive allocentric gestures, which we believe is due to the human-like, animated morphology provided by virtual arms. Second, our results suggest that robots using arms were also perceived as more likable than robots using virtual arrows, which we believe is due to that anthropomorphism and social presence. Again, we believe that while virtual arms continually draw the users' visual attention back to the robot, when virtual arrows are used, users can essentially ignore the robot generating them without any loss in performance. These findings were also observed to be highly sensitive to distance. First, the robot using virtual arms was perceived to have higher anthropomorphism when referring to objects closer to it, which we believe to be due to increased time with the animated robot in frame within the HoloLens's limited Field of View. Second, the robot using virtual arms was rated as having more likability and being more socially present when referring to objects farther from it, effects that are not yet clear how to interpret.

Finally, our results neither supported nor refuted an effect of Gesture Type on warmth or competence. We expect that these findings may in part be due to the *actual* increase in competence for robots that used virtual arrows. That is, the decreased anthropomorphism and social presence may have led these robots using arms to be perceived as more competent, but overall robots using those gestures were in fact overall *less* competent in picking out target referents than robots using virtual arrows.

3.4.3 Limitations and Motivations for Further Study. The main limitation of Study 1 was our small sample size, which, while necessary due to pandemic-related campus shutdowns, yielded unnecessarily inconclusive results in some analyses. Specifically, several analyses produced Bayes Factors between 1/3 and 3, suggesting inconclusive results neither supporting nor refuting our hypotheses, and instead suggesting the need to collect more data. While in the wake of COVID-19 many experiments are moving online [22], and while some preliminary MR-for-HRI experiments have indeed been conducted online [72], the nature of this experiment (especially with respect to physical head and eye motions to shift the Field of View of Mixed Reality) is not only ill-suited to online experimentation but also would benefit from measurement options available only in person.

Our results also revealed an interesting design challenge, in which designers should use non-ego-sensitive allocentric gestures like circles and arrows if they wish to maximize short-term task performance, but should use ego-sensitive allocentric gestures like virtual arms if they wish to maximize social dimensions likely to impact long-term task performance. We wondered, however, whether these gesture categories could be used in conjunction to achieve the best of both worlds. We thus set out to explore this in a second experiment, which we describe in the next section. Conducting a second experiment also allowed us to leverage the HoloLens 2, which has a larger Field of View.

4 STUDY 2

In the previous study, we found that, regardless of referent-robot distance, the use of virtual arrows and virtual arms differentially led to objective versus subjective benefits. The key idea behind this study is that by combining these two types of gestures, we should be able to simultaneously achieve *both* types of benefits.

4.1 Hypotheses

We thus hypothesize that a robot that uses *both* non-ego-sensitive allocentric gestures (virtual arrows) *and* ego-sensitive allocentric gestures (virtual arms) will

- **H3**: achieve better accuracy and reaction time than a robot using virtual arms alone, and
- **H4**: be perceived as more anthropomorphic and socially present (and, to a lesser extent, more likable) than a robot using virtual arrows alone.

Moreover, we believe that by using an HMD (the HoloLens 2) with a larger field of view, and by comparing across more distinct referent-robot distance conditions, we will be able to achieve greater confidence in our results.

4.2 Experiment

To investigate these hypotheses, we conducted a within-subjects human-subject experiment in which participants interacted with a robot in a mixed reality HRI context. As Study 2 directly challenges the findings and assumptions in Study 1, we only state the experiment differences below.

4.2.1 Experimental Design. We used the same general experimental task (including consideration of multiple robot-referent distance conditions) to facilitate comparison to Study 1. However, Study 2 differs in two key ways. First, a third Gesture Type condition, *both the ego-sensitive and non-ego-sensitive allocentric gestures* (in which a virtual arm reached out and pointed as an arrow appeared over target referents), was included to assess H3 and H4. But second, the two robot-referent distances used were greater—the *robot-close* condition in which the robot's target referents were approximately 1 meter from the robot and 8 meters from the human, and the *robot-distant* condition, in which the robot's target referents were approximately 6.5 meters from the robot and 2.5 meters from the human—and participants in our experiment had to identify a target referent from between five rather than three possible targets. These two changes were made in order to improve the precision of the experimental paradigm without losing comparability.

4.2.2 Experimental Apparatus. In this second study, we used the same robotic platform, a Kobuki TurtleBot 2, and the second generation of Microsoft HoloLens. Compared with Microsoft HoloLens 1 with a $30^\circ \times 17.5^\circ$ Field of View, the HoloLens 2 has a slightly larger $43^\circ \times 29^\circ$ Field of View. The same mixed reality deictic gestures were used.

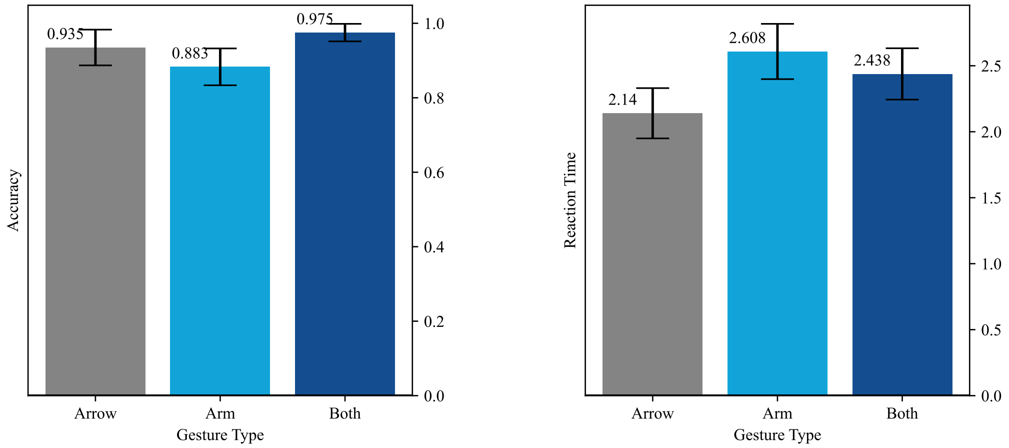
4.2.3 Procedure. Participants followed a similar procedure to Study 1. Due to COVID-19, participants were greeted virtually by an experiment supervisor through a teleconference setup on a laptop near the entry (before the arrival of each participant, experiment tools such as the HoloLens and laptop were disinfected due to COVID-19 concerns).

Participants then provided informed consent and completed a demographic survey on the laptop. Participants were then asked to watch a tutorial video instructing the participant on how to put on the HoloLens 2, open the experiment application, perform an air-tap gesture, and so forth. Once the experiment application was open, another tutorial interface within the application showed instruction text and walked participants through a sample experimental trial. After completing this tutorial, participants proceeded to the experiment. Between trials, participants took a survey as described below. After completing all six trials, participants were paid $10 and debriefed.

4.2.4 Measures. To easily compare with the results in Study 1, we used the same seven key metrics.

4.2.5 Participants. One year after Study 1, 21 participants were recruited from the Colorado School of Mines through web postings and flyers (14 male, 7 female) for an ethics-board-approved experiment. Participants ranged in age from 18 to 56 (M = 23.05, SD = 8.19). Eleven of the 21 participants had not previously engaged in any MR experiments from our lab. One participant failed to complete the experiment, leaving 20 usable data points. Objective measures (Accuracy and reaction time data) from one condition were not available for one participant, so that participant's subjective but not objective data was used in our analyses.

4.2.6 Analysis. Data analysis was performed the same as in Study 1 within a Bayesian analysis framework using the JASP 0.8.5.1 [62]. All experimental data and analysis scripts for Study 2 are available in an OSF repository (https://osf.io/asreq).



(a) **Accuracy (95% CI):** Virtual Arms and Virtual Arrows used together led to higher accuracy than Virtual Arms used alone.

(b) **Reaction Time (95% CI)**: Virtual Arrows used alone let to faster reaction times than Virtual Arms used alone.

Fig. 7. Objective results.

4.3 Results

4.3.1 Objective Results. Figure 7 summarizes our main objective results.

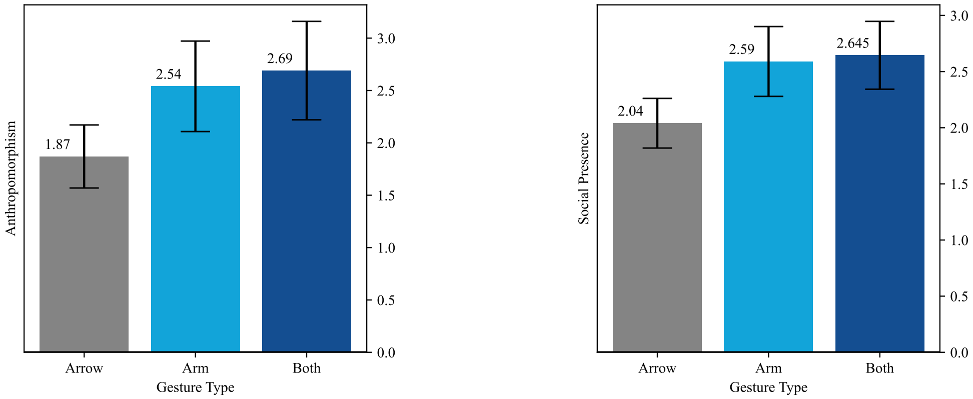
Accuracy. Our results provided moderate evidence in favor of an effect of Gesture Type on accuracy (BF 7.067) as shown in Figure 7(a), suggesting that our data was 7 times more likely to be generated under models in which Gesture Type was included than under those in which it was not. A post hoc Bayesian t-test provided very strong evidence (BF 62.590) that, specifically, when virtual arms were paired with virtual arrows, participants were more accurate in determining the intended target of those gestures (M = 0.975, SD = 0.050) than when virtual arms were used alone (M = 0.883, SD = 0.106). Anecdotal evidence was found against a difference between the use of virtual arrows alone (M = 0.935, SD = 0.103) and either virtual arms alone (BF 0.447) or arms and arrows combined (BF 0.489).

Our results also provided strong evidence in favor of an effect of Distance on accuracy (BF 11.293). Specifically, participants were generally more accurate when the referent was closer to the robot (M = 0.958, SD = 0.077) than when it was farther away (M = 0.905, SD = 0.204).

Anecdotal evidence was found against interaction between Gesture Type and Distance (BF 0.460).

Reaction Time. Our results provide very strong evidence in favor of an effect of Gesture Type on reaction time (BF 35.058) as shown in Figure 7(b). A post hoc Bayesian t-test provided extreme evidence (BF 219.542) that when virtual arrows were used by themselves, participants were faster in reacting (M = 2.140 log seconds, SD = 0.406) than when virtual arms were used by themselves (M = 2.608 log seconds, SD = 0.448). Anecdotal evidence (BF 1.682) was found in favor of a difference between the use of virtual arrows alone and arms and arrows combined (M = 2.438 log seconds, SD = 0.415), and against a difference between the use of virtual arrows alone and virtual arms alone (BF 0.788).

Our results also provided moderate evidence in favor of an effect of Distance on reaction time (BF 5.506). Specifically, participants were faster in reacting when the referent was closer to the



(a) **Anthropomorphism (95% CI):** Virtual Arms, whether used alone or when used together with Virtual Arrows, led robots to be perceived as more anthropomorphic than did Virtual Arrows used alone.

(b) **Social Presence (95% CI):** Virtual Arms, whether used alone or when used together with Virtual Arrows, led robots to be perceived as having more social presence than did Virtual Arrows used alone.

Fig. 8. Subjective results.

robot (M = 2.332 log seconds, SD = 0.490) than when it was farther away (M = 2.492 log seconds, SD = 0.348).

Moderate evidence was found *against* interaction between Gesture Type and Distance (BF 0.241).

4.3.2 Subjective Results. Figure 8 summarizes our main subjective results.

Social Presence. Our results provided extreme evidence in favor of an effect of Gesture Type on social presence (BF 12185.985) as shown in Figure 8(b). A post hoc Bayesian t-test provided extreme evidence (BF 4834.834) that when virtual arms were paired with virtual arrows, participants viewed the robots as having more social presence (M = 2.645, SD = 0.644) than when just virtual arrows were used (M = 2.040, SD = 0.473). The data also provided very strong evidence (BF 63.812) that virtual arms alone (M = 2.590, SD = 0.664) were viewed as having more social presence than virtual arrows alone. Moderate evidence (BF 0.198) was also found against a difference between pairing virtual arms and virtual arrows versus virtual arms alone.

Moderate evidence was found against an effect of Distance (BF 0.203), and anecdotal evidence was found against an interaction between Gesture Type and Distance (BF 0.444).

Anthropomorphism. Our results provided extreme evidence in favor of an effect of Gesture Type on anthropomorphism (BF 12834.920), as shown in Figure 8(a). Specifically, a post hoc Bayesian t-test provided extreme evidence (BF 420.996) that when both a virtual arrow and virtual arm were used in combination, participants found the robots to possess more anthropomorphism (M = 2.690, SD = 1.004) than when just a virtual arrow was utilized (M = 1.870, SD = 0.643); extreme evidence (BF 119.650) that virtual arms alone (M = 2.540, SD = 0.922) were viewed as having more anthropomorphism than virtual arrows alone; and moderate evidence (BF 0.110) against a difference between pairing virtual arms and virtual arrows versus virtual arms alone.

Moderate evidence was found against an effect of Distance (BF 0.207) or an interaction between Gesture Type and Distance (BF 0.179).

Likability. Moderate evidence was found against an effect of Gesture Type (BF 0.243), Distance (BF 0.216), or an interaction between the two (BF 0.230) on Likability.

Warmth. Our results provided anecdotal evidence against an effect of Gesture type on warmth (BF 0.657). A post hoc Bayesian t-test provided anecdotal evidence (BF 1.102) suggesting that when just a virtual arm was used, participants may have viewed the robot as warmer (M = 2.525, SD = 0.599) than when just virtual arrows were used (M = 2.272, SD = 0.568) but provided anecdotal evidence (BF 0.723) against a difference between virtual arrows and the use of arrows and arms combined (M = 2.479, SD = 0.753), and moderate evidence (BF 0.191) against a difference between virtual arms and the use of arrows and arms combined.

Moderate evidence was found against an effect of Distance (BF 0.244) or an interaction between Gesture Type and Distance (BF 0.261) on Warmth.

Competence. Moderate evidence was found against an effect of Gesture Type (BF 0.131) or Distance (BF 0.198) or an interaction between the two factors (BF 0.297) on Competence.

4.4 Discussion

4.4.1 Hypothesis Three.

We hypothesized that a robot that used *both* non-ego-sensitive allocentric gestures (virtual arrows) *and* ego-sensitive allocentric gestures (virtual arms) would achieve better accuracy and reaction time than a robot using virtual arms alone. Based on Study 1, we did not expect this benefit to differ between distance conditions.

Our results partially support this hypothesis. On the one hand, combining the two gesture types did indeed facilitate *accurate* referent identification. In fact, in this experiment the use of virtual arrows *only* led to demonstrably better accuracy when paired with virtual arms. That is, combining the two gestures not only led to comparable accuracy to virtual arrows but also was in fact the best Gesture Type with respect to accuracy. On the other hand, however, combining the two gesture types did *not* facilitate *fast* referent identification. While the combination of gestures did not slow reaction time as clearly as using arms alone, there was still evidence to suggest it may have yielded slower reaction time than using arrows alone. While distance was shown to impact both accuracy and reaction time, the effects of Gesture Type on those objective measures were demonstrated not to be mediated by distance.

One explanation for the results may be that while participants did not *need* to spend time watching the arm animation before selecting a referent when both gestures were used in conjunction, they chose to do so anyway, and that doing so increased accuracy by forcing them to slow down and take their time. This information about the time course of user gesture interpretation suggests future work should assess the extent to which the time course of arm animation is responsible for the enhanced accuracy and social perception observed. Another explanation could be that the response time increased when both gestures are combined due to increased need to compare and reconcile the information coming from teaching of the two visualizations.

4.4.2 Hypothesis Four.

We hypothesized that a robot that used *both* non-ego-sensitive allocentric gestures (virtual arrows) *and* ego-sensitive allocentric gestures (virtual arms) would be perceived as more anthropomorphic and socially present (and, to a lesser extent, more likable) than a robot using virtual arrows alone. Based on Study 1, we did not expect this benefit to differ between distance conditions, and we did not expect to see gesture-based differences in warmth or competence, although these measures were collected to facilitate comparison with previous work.

Our results partially supported this hypothesis. On the one hand, combining the two gesture types did lead to increased social presence and anthropomorphism comparable to the virtual arm alone, and in fact combining the two led to an even clearer difference in social presence over virtual arrows than did virtual arms alone. Moreover, the Bayes Factors found for these effects are significantly larger than those found in Study 1, despite a slightly smaller sample size (with evidence roughly 27 times stronger for an effect of gesture type on social presence and roughly twice as strong for an effect of gesture type on anthropomorphism). On the other hand, however, while we expected to find a small difference in likability based on gesture type, no differences in likability were found between any of the gestural conditions. These results help to temper the already tentative likability-based claims made in Study 1 but help to bolster the evidence that certain gesture types lead to higher levels of anthropomorphism and social presence. Finally, our results regarding warmth and competence are similar to those of Study 1, with the evidence either against an effect of Gesture Type on Warmth and Competence or inconclusive. All of these effects were shown to be invariant to distance.

4.4.3 General Discussion. Overall, our results tentatively suggest that by combining ego-sensitive allocentric gestures (virtual arms) with non-ego-sensitive allocentric gestures (virtual arrows), mixed reality robotic interaction designers can simultaneously encourage good (objective) task performance (which is important as a short-term metric) and good (subjective) social perceptions (which are important as long-term metrics), but also suggest that the benefits to accuracy of this combination of gestures (compared to using virtual arrows alone) may come at a slight reaction time cost, and that the perceptions of social presence and anthropomorphism associated with the virtual arm (both alone and when paired with virtual arrows) may not actually lead to increased likability as previously suggested. Furthermore, our results show that the insensitivity of these findings to distance was maintained even with a more clearly articulated difference between distance conditions. We would also argue that designers can be more confident in these conclusions than those presented by previous researchers, due to our use of an HMD with a larger field of view.

4.4.4 Limitations. Similar to Study 1, the main limitation of this experiment is small sample size due to COVID-19 recruiting patterns [22], although this was partially ameliorated by our within-subjects design and use of Bayesian analysis [66]. In addition, safety procedures conducted to adhere to COVID-19 guidelines (e.g., participants interacting with experimenters through a remote interface) made it difficult to identify and resolve problems arising during the experiment. In addition, our small sample was drawn from students at a small engineering university, who not only tend to be disproportionately young, male, and well educated but also are significantly more likely to have had prior experience with robotic technologies than other populations—and this is especially true for our participants, who volunteered to take part in a robotics experiment.

Another potential limitation of this work is that in the experiments presented, shadows were not shown under virtual objects. While this was not the focus of our research, previous work has shown that including these sorts of shadows can have a significant impact on users' perceptions of distance and placement. While this is highly challenging with existing technologies due to the need for light source estimation, in future work this could nevertheless be considered in replication experiments [19].

5 FUTURE WORK

In the future, one may consider the *cognitive and behavioral* impacts of mixed reality deictic gestures. While this work demonstrated benefits to objective task measures and subjective social measures to combining mixed reality deictic gestures, it is possible that this combination could be mentally taxing, especially with respect to visual perceptual load. Future work could thus explicitly assess, using self-report or neurophysiological measures, the impact of different gestures on different types of cognitive load informed by Multiple Resource Theory. In addition, work on Mixed Reality Deictic Gestures has hypothesized that objective and subjective differences between gestures may be due to differences in gaze behavior encouraged by those different gestures. Future work could thus use the Hololens 2's Eye Tracking capabilities to explicitly test this hypothesis.

Future work should also investigate the use of our selected gesture types in more dense and complex environments. One implication of the relative sparsity of a rigidly controlled experimental environment is that any communicative visualizations appearing in the environment could clearly be attributed to the single communicator in the environment, i.e., the robot. In more dense and complex environments, especially those with competing sources of visualizations and other communicators, this may not be the case. In such cases, it may be difficult for users to determine the origin of visualizations like arrows that are not obviously tied to a given communicator. Future work should thus consider both more complex environments and means for connecting visualizations like arrows back to their robot communicator, like through dashed lines, physical movement of the arrow, or other visualizations or animations.

Finally, in this work we compared ego-sensitive and non-ego-sensitive allocentric deictic gestures and their combinations. In future work, it may also be interesting to compare these variations to a use case in which a robot has a virtual arm but refuses to use it for gesturing in certain contexts, instead relying on arrow visualizations alone.

6 CONCLUSIONS

We conducted two experiments to explore the combination of two categories of mixed reality deictic gestures for armless robots: a virtual arrow positioned over a target referent (a non-ego-sensitive allocentric gesture) and a virtual arm positioned over the robot (an ego-sensitive allocentric gesture).

Our results suggest that individually, ego-sensitive gestures enable higher perceived social presence and anthropomorphism and non-ego-sensitive allocentric gestures enable faster reaction time, and that when the two gestures are used together, robots not only maintain high social presence and anthropomorphism as well as benefits (albeit not as sizeable) to reaction time but also enable uniquely improved accuracy. These results clearly suggest that when possible, these two gestures should be used in combination rather than viewed as mutually exclusive options, as their combination may indeed provide the best of both worlds.

From a robot designer's view, we make the following two context-aware recommendations for mixed reality deictic gestures:

- If a task is time-critical, an arrow should be used for pointing as it takes less time for people to react.
- If how people feel has a higher priority, the robot should use an arm for deictic gestures.

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] Adrian Bangerter and Max M. Louwerse. 2005. Focusing attention with deictic gestures and linguistic expressions. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 27.

[3] 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.

[4] Elizabeth Bates. 1976. *Language and Context: The Acquisition of Pragmatics*. Academic Press.

[5] 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.

[6] Cynthia Breazeal, Cory D. Kidd, Andrea Lockerd Thomaz, Guy Hoffman, and Matt Berlin. 2005. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 708–713.

[7] 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.

[8] Judee K. Burgoon and Gregory D. Hoobler. 1994. Nonverbal signals. *Handbook of Interpersonal Communication* 2 (1994), 229–285.

[9] Rehj Cantrell, Paul Schermerhorn, and Matthias Scheutz. 2011. Learning actions from human-robot dialogues. In *Proceeding of the International Symposium on Robot-Human Interactive Communication (RO-MAN).* IEEE, 125–130.

[10] 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.

[11] 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.

[12] Ravi Teja Chadalavada, Henrik Andreasson, Robert Krug, and Achim J. Lilienthal. 2015. That's on my mind! Robot to human intention communication through on-board projection on shared floor space. In *2015 European Conference on Mobile Robots (ECMR'15).* IEEE, 1–6.

[13] Eve V. Clark and C. J. Sengul. 1978. Strategies in the acquisition of deixis. *Journal of Child Language* 5, 3 (1978), 457–475.

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

[15] Susan Wagner Cook, Terina Kuangyi Yip, and Susan Goldin-Meadow. 2012. Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes* 27, 4 (2012), 594–610.

[16] Martin J. Crowder. 2017. *Analysis of Repeated Measures.* Routledge.

[17] Antonella De Angeli, Walter Gerbino, Giulia Cassano, and Daniela Petrelli. 1998. Visual display, pointing, and natural language: The power of multimodal interaction. In *Proceedings of the Working Conference on Advanced Visual Interfaces.* 164–173.

[18] Ewart J. De Visser, Samuel S. Monfort, Ryan McKendrick, Melissa A. B. Smith, Patrick E. McKnight, Frank Krueger, and Raja Parasuraman. 2016. Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied* 22, 3 (2016), 331.

[19] 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'17).* IEEE, 111–122.

[20] 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.

[21] Brian R. Duffy. 2002. Anthropomorphism and robotics. *Robotics and Autonomous Systems* 20 (2002).

[22] David Feil-Seifer, Kerstin S. Haring, Silvia Rossi, Alan R. Wagner, and Tom Williams. 2020. Where to next? The impact of COVID-19 on human-robot interaction research. *ACM Transactions on Human-Robot Interaction (THRI)* 10, 1 (2020), 1–7.

[23] 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.

[24] Susan T. Fiske, Amy J. C. Cuddy, and Peter Glick. 2007. Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences* 11, 2 (2007), 77–83.

[25] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. 2003. A survey of socially interactive robots. *Robotics and Autonomous Systems* 42, 3–4 (2003), 143–166.

[26] 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.

[27] 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.

[28] Ipek Goktan, Karen Ly, Thomas Roy Groechel, and Maja Mataric. 2022. Augmented reality appendages for robots: Design considerations and recommendations for maximizing social and functional perception. In *5th International Workshop on Virtual, Augmented, and Mixed Reality for HRI.*

[29] Susan Goldin-Meadow. 1999. The role of gesture in communication and thinking. *Trends in Cognitive Sciences* 3, 11 (1999), 419–429.

[30] Thomas Groechel, Zhonghao Shi, Roxanna Pakkar, and Maja J. Matarić. 2019. Using socially expressive mixed reality arms for enhancing low-expressivity robots. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN'19).* IEEE, 1–8.

[31] Thomas R. Groechel, Michael E. Walker, Christine T. Chang, Eric Rosen, and Jessica Zosa Forde. 2021. TOKCS: Tool for organizing key characteristics of VAM-HRI systems. *arXiv preprint arXiv:2108.03477* (2021).

[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] Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. De Visser, and Raja Parasuraman. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors* 53, 5 (2011), 517–527.

[34] Simon Harrison. 2011. The creation and implementation of a gesture code for factory communication. In *Gesture and Speech in Interaction (GESPIN'11)*.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] Stephen C. Levinson. 2004. 5 Deixis. *Handbook of Pragmatics* (2004), 97.

[40] Matthew Lombard and Theresa Ditton. 1997. At the heart of it all: The concept of presence. *Journal of Computer-mediated Communication* 3, 2 (1997), JCMC321.

[41] Bertram F. Malle and Matthias Scheutz. 2014. Moral competence in social robots. In *2014 IEEE International Symposium on Ethics in Science, Technology and Engineering*. IEEE, 1–6.

[42] William Marslen-Wilson, Elena Levy, and Lorraine K. Tyler. 1982. Producing interpretable discourse: The establishment and maintenance of reference. In *Proceeding of the Speech, Place, and Action. Studies in Deixis and Related Topics*, Robert J. Jarvella and Wolfgang Klein (Eds.). John Wiley & Sons Ltd, Chichester, 339–378.

[43] S. Mathôt. 2017. Bayes Like a Baws: Interpreting Bayesian Repeated Measures in JASP [Blog Post]. https://www.cogsci.nl/blog/interpreting-bayesian-repeated-measures-in-jasp.

[44] Nikolaos Mavridis. 2015. A review of verbal and non-verbal human–robot interactive communication. *Robotics and Autonomous Systems* 63 (2015), 22–35.

[45] R. D. Morey and J. N. Rouder. 2014. BayesFactor (Version 0.9.9).

[46] 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.

[47] 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.

[48] Yusuke Okuno, Takayuki Kanda, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. 2009. Providing route directions: Design of robot's utterance, gesture, and timing. In *2009 4th ACM/IEEE International Conference on Human-robot Interaction (HRI'09)*. IEEE, 53–60.

[49] Raedy Ping and Susan Goldin-Meadow. 2010. Gesturing saves cognitive resources when talking about nonpresent objects. *Cognitive Science* 34, 4 (2010), 602–619.

[50] Christopher Reardon, Kevin Lee, and Jonathan Fink. 2018. Come see this! Augmented reality to enable human-robot cooperative search. In *2018 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR'18)*. IEEE, 1–7.

[51] Laurel D. Riek, Philip C. Paul, and Peter Robinson. 2010. When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry. *Journal on Multimodal User Interfaces* 3, 1–2 (2010), 99–108.

[52] Paul Robinette, Ayanna M. Howard, and Alan R. Wagner. 2015. Timing is key for robot trust repair. In *International Conference on Social Robotics*. Springer, 574–583.

[53] 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.

[54] 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.

[55] 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.

[56] Allison Sauppé and Bilge Mutlu. 2014. Robot deictics: How gesture and context shape referential communication. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI'14)*. IEEE, 342–349.

[57] 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. *arXiv preprint arXiv:2008.05799* (2020).

[58] Matthias Scheutz, Paul Schermerhorn, James Kramer, and David Anderson. 2007. First steps toward natural human-like HRI. *Autonomous Robots* 22, 4 (2007), 411–423.

[59] Elena Sibirtseva, Dimosthenis Kontogiorgos, Olov Nykvist, Hakan Karaoguz, Iolanda Leite, Joakim Gustafson, and Danica Kragic. 2018. A comparison of visualisation methods for disambiguating verbal requests in human-robot interaction. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'18)*. IEEE, 43–50.

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

[61] Daniel Szafir, Bilge Mutlu, and Terrence Fong. 2015. Communicating directionality in flying robots. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI'15)*. IEEE, 19–26.

[62] JASP Team. 2018. JASP (Version 0.8.5.1)[Computer Software].

[63] Stefanie Tellex, Nakul Gopalan, Hadas Kress-Gazit, and Cynthia Matuszek. 2020. Robots that use language. *Annual Review of Control, Robotics, and Autonomous Systems* 3, 1 (2020).

[64] Nhan Tran, Trevor Grant, Thao Phung, Leanne Hirshfield, Christopher D. Wickens, and Tom Williams. 2021. Robot-generated mixed reality gestures improve human-robot interaction. In *Proceedings of the International Conference on Social Robotics*.

[65] Nhan Tran, Kai Mizuno, Trevor Grant, Thao Phung, Leanne Hirshfield, and Tom Williams. 2019. Exploring mixed reality robot communication under different types of mental workload. In *Proceedings of the 3rd International Workshop on Virtual, Augmented, and Mixed Reality for HRI*.

[66] Rens Van De Schoot, Joris J. Broere, Koen H. Perryck, Mariëlle Zondervan-Zwijnenburg, and Nancy E. Van Loey. 2015. Analyzing small data sets using Bayesian estimation: The case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *European Journal of Psychotraumatology* 6, 1 (2015), 25216.

[67] E. J. Wagenmakers, J. Love, M. Marsman, T. Jamil, A. Ly, and J. Verhagen. 2018. Bayesian inference for psychology, Part II: Example applications with JASP. *Psychonomic Bulletin and Review* 25, 1 (2018), 35–57.

[68] 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).

[69] Atsushi Watanabe, Tetsushi Ikeda, Yoichi Morales, Kazuhiko Shinozawa, Takahiro Miyashita, and Norihiro Hagita. 2015. Communicating robotic navigational intentions. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'15)*. IEEE, 5763–5769.

[70] 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.

[71] Thomas Weng, Leah Perlmutter, Stefanos Nikolaidis, Siddhartha Srinivasa, and Maya Cakmak. 2019. Robot object referencing through legible situated projections. In *2019 International Conference on Robotics and Automation (ICRA'19)*. IEEE, 8004–8010.

[72] 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*.

[73] 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*.

[74] 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 *International Conference on Human-Computer Interaction*. Springer, 178–198.

[75] Tom Williams and Matthias Scheutz. 2016. A framework for resolving open-world referential expressions in distributed heterogeneous knowledge bases. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*.

[76] Tom Williams and Matthias Scheutz. 2019. Reference in robotics: A givenness hierarchy theoretic approach. In *Proceedings of the Oxford Handbook of Reference*, Jeanette Gundel and Barbara Abbott (Eds.). Oxford University Press, 457–474.

[77] Tom Williams, Daniel Szafir, and Tathagata Chakraborti. 2019. The reality-virtuality interaction cube. In *Proceedings of the 2nd International Workshop on Virtual, Augmented, and Mixed Reality for HRI*.

[78] 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.

[79] 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*.

[80] 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*.

Received 1 March 2022; revised 25 May 2022; accepted 1 September 2022