



Investigating the Potential Effectiveness of Allocentric Mixed Reality Deictic Gesture

Tom Williams^(✉), Matthew Bussing, Sebastian Cabrol, Ian Lau,
Elizabeth Boyle, and Nhan Tran

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

Abstract. Mixed reality technologies offer interactive robots many new ways to communicate their beliefs, desires, and intentions to human teammates. In previous work, we identified several categories of visualizations that when displayed to users through mixed reality technologies serve the same role as traditional deictic gestures (e.g., pointing). In this work, we experimentally investigate the potential utility of one of these categories, allocentric gestures, in which circles or arrows are rendered to enable human teammates to pick out the robot's target referents. Specifically, through two human subject experiments, we examine the objective and subjective performance of such gestures alone as compared to language alone and the combination of language and allocentric gesture. Our results suggest that allocentric gestures are more effective than language alone, but to maintain high robot likability allocentric gestures should be used to complement rather than replace complex referring expressions.

Keywords: Mixed reality · Augmented reality · Deixis ·
Natural language generation · Human-robot interaction

1 Introduction

Robots designed to interact with humans must be able to do so in a way that is natural, efficient, and human-like. This is especially important for many of the domains that the field of Human-Robot Interaction (HRI) is interested in, such as search-and-rescue, in which users (e.g., search-and-rescue victims) may never have interacted with a robot before, may be physically unable to control it through some other means, and may not have the time or willingness to undergo training to learn to operate the robot; or in which other users (e.g., search-and-rescue operators) may not have the cognitive bandwidth to dedicate to direct operation. Accordingly, researchers are increasingly turning to *natural language* as an intuitive and flexible modality for controlling and interacting with robots.

All authors are with the MIRRORLab (mirrorlab.mines.edu).

© Springer Nature Switzerland AG 2019

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2019, LNCS 11575, pp. 178–198, 2019.

https://doi.org/10.1007/978-3-030-21565-1_12

When humans verbally communicate, they typically accompany their utterances with various physical gestures. When picking out objects, people, and locations in the environment, humans typically use deictic gestures, such as pointing, to quickly draw their interlocutors' attention to their intended target without the use of overly complex description. Indeed, deictic gesture in particular is one of the most important communicative faculties available to humans, and one of the earliest arising communicative strategies in human development. Deictic gestures not only allow speakers to speak more concisely, but they lighten speakers' cognitive load [1] and working memory load [2]. Moreover, deictic gestures facilitate listeners' comprehension by amplifying semantic content [3] and shifting their attention [4], which in turn facilitates reference resolution [5] and helps to establish joint attention [6]. Indeed, in many situated contexts, it is difficult to effectively communicate *without* the use of deictic gesture.

Accordingly, HRI researchers have explored how robots might generate deictic gestures to allow robots to reap these same benefits. And while this previous work has largely been focused on pointing, HRI researcher have also studied how robots might generate many other forms of deictic gesture, such as presenting, exhibiting, touching, grouping, and sweeping, and have studied how these different forms of gesture are differentially perceived by humans [7].

In our work, we are exploring robots' use of deictic gesture beyond even this wide variety of forms, to examine entirely new classes of deictic gestures enabled by recent technological developments. Specifically, we are interested in how new Augmented and mixed reality technologies can be used to enable mixed reality Deictic Gestures: new types of deictic gestures visualized in mixed reality environments. These new forms of gesture may replace or complement traditional (physical) deictic gestures in contexts where those gestures are impossible, e.g., when a robot lacks arms with which to gesture; or in contexts in which physical gestures are possible but suboptimal, e.g., when a robot is not colocated with their human interlocutor, or in environments in which traditional gestures would be difficult to see, such as dark and dusty subterranean environments.

In Sect. 2, we briefly survey previous work on human and robot use of deictic gesture, as well as of recent work at the intersection of augmented and mixed reality and HRI, including the limited set of work previously exploring mixed reality deictic gesture for HRI. In Sects. 3 and 4, we then describe two human subject experiments designed to provide a preliminary investigation of the effectiveness and human perception of mixed reality deictic gesture, in which we assess human perceptions of videos simulating the display of one category of mixed reality deictic gestures, *allocentric gestures*. In Sect. 4.3, we then introduce preliminary work we have performed towards enabling robot generation of allocentric gestures. Finally, in Sect. 6, we conclude with several possible directions for future work.

2 Related Work

2.1 Human Deictic Gesture

Deixis is one of the most crucial forms of human-human communications [8,9], as well as one of the oldest, both anthropologically and developmentally. Humans point while speaking even from infancy, with deictic gesture beginning around 9–12 months [10], and general deictic reference mastered around age 4 [11]. Deictic gestures have been shown to be a powerful technique for language learners, as they allow speakers to communicate intended referents before being able to do so in language, just as other types of gestures help speakers to communicate their intended sense or meaning when they otherwise lack the words to do so. Indeed, developmental changes in deictic gestural capabilities in humans are a strong predictor of changes in language development [12].

In addition, long past infancy, humans continue to rely on deictic gesture as a core communicative capability, as its attention-direction presents an efficient and workload-reducing referential strategy in complex environments, far beyond that of purely verbal reference [13–17], and as deictic gesture allows for communication in environments in which verbal communication would be difficult or impossible, such as in noisy factory environments [18]. Accordingly, it is no surprise that Human-Robot Interaction researchers have sought to enable this effective and natural communication strategy in robots.

2.2 Robot Deictic Gesture

There is widespread evidence for the effectiveness of robots’ use of physical deictic gesture: studies have shown that robots’ use of deictic gesture is effective at shifting attention in the same way as is humans’ use of deictic gesture [19], and that robots’ use of deictic gesture improves both subsequent human recall and human-robot rapport [20]. This effectiveness has been demonstrated across different contextual scales as well, including gestures to nearby objects on a tabletop [21], gestures to larger regions of space between the robot and its interlocutor [22], and gesture to large-scale spatial locations during direction-giving [23]. Furthermore, this effectiveness has shown to be especially true when gestures are generated in socially appropriate ways [24].

Research has also shown that robots’ use of deictic gesture is especially effective when paired with other nonverbal signaling mechanisms [25], such as *deictic gaze*, in which a robot (actually or ostensibly) shifts its gaze towards its intended referent [22,26,27], and that this is especially effective when gaze and gesture are appropriately coordinated [28]. These findings have motivated a variety of technical approaches to deictic gesture generation [29–31], as well as a number of approaches for integrating gesture generation with natural language generation [32] (see also [33–35]).

Of particular interest is the work of Sauppé and Mutlu [7]. Building off the work of Clark, who showed that humans use many deictic gestures beyond pointing [36], Sauppé and Mutlu explored a selection of robotic deictic gestures: pointing, presenting, exhibiting, touching, grouping, and sweeping. Sauppé and Mutlu

were especially interested in how these categories differed in both effectiveness and perceived naturality, and how different contextual factors, such as the density of candidate referents, the number of fully ambiguous distractors for the referent, and the distance of the referent from the referrer. As we will describe, the set of questions we are interested in investigating both in this work and in future work has a number of parallels with those of interest to Sauppé and Mutlu, and accordingly, as we will also describe, the experiment presented in this paper was designed with careful attention to Sauppé and Mutlu’s design.

2.3 Augmented Reality for HRI

Research on augmented and mixed reality have been steadily progressing over the past several decades [37–39], and there have been a number of papers over the past twenty-five years highlighting the advantages of leveraging augmented reality (AR) technologies to facilitate human-robot interactions [40, 41]. Augmented and mixed reality technologies lies primarily in two areas: (1) their potential to increase the flexibility of users’ control over robots through visualizations of robot-controlling interface elements, and (2) their potential to increase the expressivity of users’ view into those robots’ internal states through visualizations that reflect information from the robot’s internal state [42]. Historically, however, there has been surprisingly little research on this topic.

Recently, this has changed, with research at the intersection of these fields beginning to dramatically increase [43, 44], with approaches being presented that use AR for robot design [45], calibration [46], and training [47], and for communicating robots’ perspectives [48], intentions [49, 50] and trajectories [51–53]. Most relevant to this paper are recent works on aligning human and robot perspective to facilitate communication. Amor et al. project instructions and highlight task-relevant objects, but with no language generation, and with visualizations cast as part of the environment, rather than as the robot’s communication [54]. Sibirtseva et al. present an approach in which, as a human describes a referent, the robot’s distribution over intended referents is visualized by circling remaining reference candidates in the user’s AR Head-Mounted Display (HMD) [55] (cp. [56]). The visualizations used in this work are cast as being from the robot’s perspective, but this is passive backchannel communication rather than active communication of the robot’s intentions. Also of interest is recent work from Reardon et al., in which a robot draws the trajectory a human should take into their HMD, highlighting the intended target [57]. This work takes a more active communication approach than the work of Sibirtseva et al., but does not involve robot language generation.

To explore the use of AR in active, linguistic, robotic communication, we have presented a framework for categorizing deictic gestures available in mixed reality human-robot interactions, including both traditional physical gestures and virtual deictic annotations (categorized into allocentric gestures (e.g., circling a target referent in a user’s AR HMD), perspective-free gestures (e.g., projecting a circle around a target referent on the floor of the shared environment), ego-sensitive allocentric gestures (e.g., pointing to a target referent using a simulated

arm rendered in a user’s AR HMD), and ego-sensitive perspective-free gestures (e.g., projecting a line from the robot to its target on the floor of the shared environment)), as well as combinations of different forms of mixed reality deictic gesture [58, 59]. Crucially, we argue that not only do tradeoffs exist between different forms of mixed reality deictic gestures (including differences in privacy, cost, and legibility), but moreover, that there are a number of contextual factors that dictate the circumstances in which mixed reality deictic gestures become especially valuable, including teammate workload, auditory and visual perceptual load, and so forth [60].

This framework is especially valuable for our research as, in conjunction with the work of Sauppé and Mutlu [7], it suggests concrete hypotheses regarding the effectiveness and perception of mixed reality deictic gestures in different contexts, allowing us to empirically investigate whether mixed reality deictic gestures have the same communicative benefits as physical gestures, and how those benefits differ according to context. In this paper, we present a set of such hypotheses, and describe two human subject experiments designed to investigate them [61, 62].

3 Experiment 1

Our first experiment [61] investigated the combination of language and mixed reality deictic gesture for robot communication. Here, participants viewed videos of a robot referring to 12 objects within a visual scene. This IRB-approved experiment was designed so as to follow the within-subjects paradigm used in the seminal evaluation of physical robot gesture presented by Sauppé and Mutlu [7].

3.1 Experimental Design

Interaction Design: Our first independent variable was *communication style*. For one-third of the objects, the robot used *complex reference* alone, generating an expression of the form “Look at that {color} {shape}” (e.g. “Look at that red cube”). These utterances followed a common fixed-length form even if it was not fully disambiguating, in hopes of better studying reaction time. For another third of the objects, the robot used a mixed reality deictic gesture, drawing a circle around the target and stating “Look at that” (cp. the gestural conditions used by Sauppé and Mutlu). For the remaining objects, the robot used both complex reference *and* mixed reality deictic gesture, circling the target and then generating a complex reference as described above (cp. the gestural and fully articulated conditions used by Sauppé and Mutlu).

Environment Design: The experimental environment contained a Kobuki robot positioned behind an array of eighteen blocks, of four shapes (cubes, triangles, cylinders, towers) and four colors (red, yellow, green, blue), evenly spaced in four rows. Specifically, there were six unique blocks and six pairs of non-unique blocks (a difference of *inherent ambiguity*), evenly split between the front and rear rows (a difference of *distance*), and distributed as uniformly as possible

according to color and shape (as shown in Fig. 1). This sought to simultaneously capture multiple environmental dimensions previously determined by Sauppé and Mutlu to affect the accuracy and perceived effectiveness of reference: *ambiguity* and *distance from referrer* while controlling for the other dimensions previously investigated by Sauppé and Mutlu (object clustering, visibility, and noise). Our second and third independent variables were thus referent ambiguity and referent distance¹, yielding a total of twelve ($3 \times 2 \times 2$) experimental conditions.



Fig. 1. Task environment, with simulated AR visualization (Color figure online)

3.2 Procedure

Participants were recruited online using Amazon’s Mechanical Turk platform, and directed towards a psiTurk experimental environment [63]. After providing informed consent and providing demographic information, participants were instructed that they would watch a series of videos in which a robot described and/or visually gestured towards a target object by drawing a circle around it. They were told that they should click on the object that was being described as soon as they had identified it. Participants were then assigned to one of twelve conditions each corresponding to a different video order determined through a counterbalanced Latin Square array. Participants then watched twelve videos, each corresponding with a different experimental condition. When mixed reality deictic gesture was used in a video, gesture onset began 660 ms before speech onset, based on the gestural timing model presented by Huang and Mutlu [64] and leveraged by Sauppé and Mutlu [7]. Clicking on any object within a video sent the participant to a survey page in which they were asked to assess the effectiveness of the robot’s speech and gesture and the likability of the robot, using the measures described below. Upon answering these survey questions, participants were allowed to proceed to the next video in the series. All videos were six seconds in length, including padding before and after the robot’s utterance.

¹ We did not expect to see effects of distance, but included distance as an independent variable so that we can use an identical experimental design in future experiments in which we will use other types of gestures, e.g., pointing gestures generated with real or simulated arms, for which we *would* expect to see a potential difference.

3.3 Hypotheses

In this experiment, we examined four core hypotheses:

- H1.** We hypothesized that participants would have worse accuracy in identifying the robot’s target referent only when ambiguous complex noun phrases were used without an associated mixed reality deictic gesture (i.e., in the complex reference condition for targets with inherent ambiguity).
- H2.** We hypothesized (**H2.1**) that the speed at which participants would be able to identify the robot’s target referent would be better when mixed reality deictic gesture was used, as it would allow target referents to be disambiguated even before speech began, and (**H2.2**) that this reaction time would increase when a reference was ambiguous.
- H3.** We hypothesized (**H3.1**) that participants would perceive the robot to be more effective when mixed reality deictic gesture was used, especially (**H3.2**) when used in combination with complex reference, and (**H3.3**) when the target referent was ambiguous.
- H4.** We hypothesized that the robot’s likability would correlate with its effectiveness, and accordingly, that (**H4.1**) perceived likability would be higher when mixed reality deictic gesture was used, (**H4.2**) especially in combination with complex reference, and (**H4.3**) for ambiguous targets.

3.4 Measures

To assess these hypotheses, objective and subjective measures were used.

Accuracy: An objective measure of *accuracy* was gathered by recording which item in each scene participants clicked on, and determining whether or not this was in fact the object intended by the robot.

Reaction Time: An objective measure of *reaction time* was gathered by recording time stamps at the moment each video phase began (i.e., when the page loaded) and ended (i.e., when an object was clicked on).

Effectiveness: A subjective measure of robot *effectiveness* was gathered using a version of the Gesture Perception Scale [7] modified to make reference to mixed reality deictic gesture rather than simply gesture [61]. Each participants’ scores for each video were then transformed to a range of 0–100 and averaged. A reliability analysis indicated that the internal reliability of this scale was very high for our experiment, with Cronbach’s $\alpha = 0.955$.

Likability: A subjective measure of robot *likability* was gathered using the Godspeed II Likability scale [65]. Our modified version asked participants to rate their perception of the robot along each dimension by clicking a point anywhere along a five-point Likert-type scale. Each participants’ scores for each video were transformed to a range of 0–100 and averaged. A reliability analysis indicated very high internal reliability (Cronbach’s $\alpha = 0.963$).

3.5 Participants

50 participants were recruited from Amazon Mechanical Turk (19 F, 31 M). Participants ranged in age from 19 to 69 ($M = 39.07$, $SD = 11.35$). None had participated in any previous studies from our laboratory under the account used.

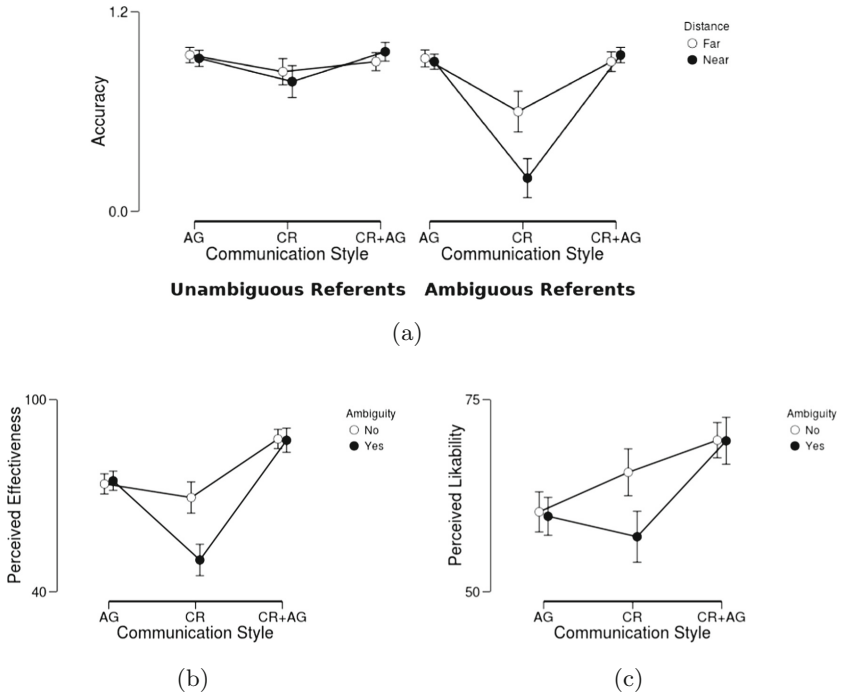


Fig. 2. (a) Effect of communication style (augmented gesture (AG), vs complex reference (CR) vs both (CR+AG)), referent ambiguity and referent distance on participant accuracy. (b) Effect of communication style (augmented gesture (AG), vs complex reference (CR) vs both (CR+AG)) and referent ambiguity on perceived effectiveness. (c) Effect of communication style (augmented gesture (AG), vs complex reference (CR) vs both (CR+AG)) and referent ambiguity on likability

3.6 Analysis

Data analysis was performed within a Bayesian analysis framework using the JASP 0.8.5.1 [66] software package, using the default settings as justified by Wagenmakers et al. [67]. All data files are available at tinyurl.com/hri19data. For each measure, a repeated measures analysis of variance (RM-ANOVA) [68–70] was performed, using communication style, ambiguity, and distance as random factors. Baws factors [71] were then computed for each candidate main effect and interaction, indicating (in the form of a Bayes Factor) for that effect the evidence weight of all candidate models including that effect compared to the evidence

weight of all candidate models not including that effect. When sufficient evidence was found in favor of a main effect of communication style (a three-level factor), the results were further analyzed using a post-hoc Bayesian t-test [72, 73] with a default Cauchy prior (center = 0, $r = \frac{\sqrt{2}}{2} = 0.707$).

3.7 Results

Accuracy: We hypothesized (**H1**) that accuracy would only drop when ambiguous complex noun phrases were used without an associated mixed reality deictic gesture (i.e., in the complex reference condition). Our results provided extreme evidence in favor of an effect of communication style (Bf 5.626e28)² and ambiguity (Bf 2.380e7), and for interactions between communication style and both ambiguity (Bf 1.521e13) and distance (Bf 44577.358). In addition, strong evidence was found in favor of a three-way interaction (22.183).

Main Effect: Communication Style: Post-hoc analysis provided extreme evidence for differences in accuracy, specifically between the complex reference condition ($M = 0.605$, $SD = 0.49$) and both the mixed reality deictic gesture condition ($M = 0.92$, $SD = 0.272$) (Bf 1.129e15) and the complex reference + mixed reality deictic gesture condition ($M = 0.925$, $SD = 0.264$) (Bf 4.728e13). This suggests that the use of complex reference by itself was significantly less effective than mixed reality deictic gesture.

Main Effect: Ambiguity: Our results also suggest that accuracy was worse when the robot referred to an ambiguous referent ($M = 0.743$, $SD = 0.438$) than when it referred to an unambiguous referent ($M = 0.89$, $SD = 0.313$).

Interaction: Communication Style and Ambiguity: These results are clarified by the interaction found between communication style and ambiguity: performance was only *much* worse when using ambiguous complex references without an associated gesture. This confirms hypothesis **H1**.

Interaction: Communication Style and Distance: Our results demonstrate that when a target referent was close to the robot, using a complex reference alone significantly harmed performance more than when the referent was far away.

Interaction: Communication Style, Ambiguity, and Distance: This effect is further clarified through the three-way interaction, which shows that performance drops only occurred when the reference was ambiguous, as shown in Fig. 2a.

Reaction Time: We hypothesized (**H2.1**) that reaction time would drop when mixed reality deictic gesture was used, as it would allow target referents to be disambiguated even before speech began, and (**H2.2**) that reaction time would increase when a reference was ambiguous. No results were found in favor of our

² Bayes Factors above 100 indicate extreme evidence in favor of a hypothesis [74]. Here, for example, our Baws Factor Bf of 5.626e28 suggests that our data were 5.626e28 times more likely to be generated under models in which communication style is included than under those in which it is not.

hypotheses: in fact, our analysis provided strong evidence against a main effect of ambiguity or any interaction effect. Median reaction time was 7.7 s.

Effectiveness: We hypothesized (**H3.1**) that perceived effectiveness would be higher when mixed reality deictic gesture was used, especially (**H3.2**) when used in combination with complex reference, and (**H3.3**) when the target referent was ambiguous. Our results provided extreme evidence in favor of main effects of communication style (Bf 1.601e36) and ambiguity (Bf 216.516), and for an interaction between communication style and ambiguity (Bf 1.04e6).

Main Effect: Communication Style: Post-hoc analysis provided extreme evidence in favor of a difference in perceived effectiveness between communication styles (mixed reality deictic gesture (M = 74.17, SD = 23.59) vs. complex reference (M = 59.67, SD = 27.30) (Bf 2.038e7); mixed reality deictic gesture vs. complex reference + mixed reality deictic gesture (M = 87.50, SD = 17.08) (Bf 1.462e10); complex reference vs complex reference + mixed reality deictic gesture (Bf 1.581e23)). Specifically, our results show a strong perceived ordering in effectiveness: complex reference < mixed reality deictic gesture < complex reference + mixed reality deictic gesture. This confirms hypotheses **H3.1** and **H3.2**.

Main Effect: Ambiguity: In addition, our results showed that robots were perceived as less effective when describing ambiguous referents (M = 70.63, SD = 26.98) than when describing unambiguous referents (M = 76.93, SD = 23.92).

Interaction: Communication Style and Ambiguity: These results are clarified by examining the observed interaction between communication style and ambiguity, which suggests that while the robot was perceived as less effective when using complex reference alone even when the referent was unambiguous, the robot was perceived as *much* less effective when using complex reference alone to describe ambiguous targets, as seen in Fig. 2b. This confirms hypothesis **H3.3**.

Likability: We hypothesized that robots' perceived likability would correlate with their perceived effectiveness, and that (**H4.1**) perceived likability would be higher when mixed reality deictic gesture was used, (**H4.2**) especially in combination with complex reference, and (**H4.3**) when the target referent was ambiguous. Our results provided extreme evidence in favor of a main effect of communication (Bf 5.986e9), and moderate evidence in favor of an effect of ambiguity (Bf 3.088) or its interaction with communication (Bf 7.985).

Main Effect: Communication Style: Post-hoc analysis provided extreme evidence in favor of a difference in likability between the use of complex reference *and* mixed reality deictic gesture (M = 69.68, SD = 19.27) and the use of *either* complex reference (M = 61.35, SD = 22.40) (Bf 81289.052) *or* mixed reality deictic gesture (M = 60.11, SD = 19.64) (Bf 9.940e7). This suggests that participants much more strongly liked the robot when it used both communication styles in combination, confirming hypothesis **H4.1**.

Main Effect: Ambiguity: Our results suggested that participants liked the robot less when it referred to ambiguous referents.

Interaction: Communication Style and Ambiguity: This interaction effect suggested that when the robot’s target referent was unambiguous, participants exhibited a likability ordering: mixed reality deictic gesture < complex reference < mixed reality deictic gesture + complex reference; but when the robot’s target referent ambiguous, participants particularly disliked the use of complex reference alone (which is unsurprising given that in such cases complex reference alone did not allow the target to be properly disambiguated). These findings, as seen in Fig. 2c, confirming hypotheses **H4.3** and partially supporting **H4.2**.

3.8 Discussion

Our results suggest that mixed reality deictic gestures may be an accurate, likable, and effective communication strategy for human-robot interaction, much the same as traditional physical deictic gestures. In this section, we will discuss these results in detail, and leverage them to produce design guidelines for enabling mixed reality deictic gestures.

Objective Effectiveness of Mixed Reality Deictic Gesture: Our first and second hypotheses considered the objective effectiveness of mixed reality deictic gestures. Specifically, we hypothesized (**H2.1**) that mixed reality deictic gestures would facilitate faster human reference resolution, especially in the case of ambiguous referents (**H2.2**) – for which referents we also hypothesized that mixed reality deictic gesture would enable increased accuracy (**H1**). Our results did indeed suggest that participants had better accuracy in selecting ambiguous referents when mixed reality deictic gestures were used, and especially when referents were ambiguous (supporting **H1**). This is not particularly surprising, as when complex reference alone was used to refer to otherwise ambiguous referents, the specific descriptions we used were not themselves sufficient to disambiguate those referents. Specifically, to appropriately control language complexity, all instances of complex reference took the form “Look at that {color} {shape}”. When a referent was ambiguous (i.e., there were more than one object of that color and shape), clearly this expression itself was still ambiguous. As we will describe below, in our second experiment we sought to instead use a complex reference condition that more fully aligned with the “fully articulated” baseline used by Sauppé and Mutlu [7], which sacrifices control over linguistic complexity for assurance of complete disambiguation. This draws an interesting contrast with Sauppé and Mutlu’s experiment, in which the fully articulated baseline was fully disambiguating, but the majority of the deictic gestures examined were *not*; the opposite pattern as observed in our own experimental design.

But while our first hypothesis was supported, no effects on reaction time were observed, thus failing to support **H2**. As median reaction time was 7.7 s for videos that were around 5–6 s in length, this suggests that participants nearly uniformly waited until videos completed before selecting their targets, and were not hindered by ambiguity. We expected that despite our instructions to click on target referents as soon as they were identified, participants may simply

not have been aware of the ability or benefit of doing so. As we will describe below, we sought to address this concern in our second experiment. In future work, it could also be interesting to gain an even more fine-grained measure of how mixed reality deictic reference affects reaction time in complex, multi-entity reference, using eye-tracking techniques such as those employed in Visual World-paradigmatic experiments [75].

In addition, we found a surprising interaction between communication style and distance. We believe that this finding may best be explained by imagining an attentional cone extending in front of the robot. Several theories of qualitative spatial reference (e.g., Ternary Point Configuration Calculus [76]) consider one entity to be “in front” of another if it falls within just such a cone. Our results suggest that when participants had to choose between options that had not been fully disambiguated, they were biased towards options that could be considered to be “in front” of the robot because they fell within that cone. Because of the conic nature of this region, all objects far from the robot may have been considered “in front” of the robot, yielding no bias for any particular distant object, whereas only some of the objects close to the robot would have been considered “in front” of it, yielding a bias towards those objects. This led to poor accuracy in cases of ambiguity where the “true” target referent did not fall within that attentional cone. We would also note that our experimental design uniquely enabled us to identify this interaction; no such interaction was observed by Sauppé and Mutlu because their experimental design did not allow distance and ambiguity to be simultaneously investigated. That being said, as we will describe below, in our second experiment, we sought to remove the need for participants to occasionally select between not-fully-disambiguated referents, trading ability to investigate this effect for the ability to better capture the overall potential for impact for mixed reality deictic gesture.

Subjective Perceptions of Mixed Reality Deictic Gesture: Our final hypotheses considered the subjective perception of mixed reality deictic gestures. Specifically, we hypothesized (**H3.1/H4.1**) that participants would perceive the robot to be more effective and likable when mixed reality deictic gesture was used, especially when used in combination with complex reference (**H3.2/4.2**), and when used to refer to an otherwise ambiguous referent (**H3.3/4.3**).

Our results supported all of these hypotheses, with the possible exception of **H4.2**: when the target referent would *not* have been otherwise ambiguous, participants actually reported liking the robot more when complex reference alone was used than for mixed reality deictic gesture alone (accompanied only by a minimally articulated verbal reference). This serves to emphasize that, like physical gesture, mixed reality deictic gesture should be used to supplement rather than replace natural language (excepting extreme circumstances). However, clearly these differences may be exaggerated by the same features of our complex references that potentially exaggerated the accuracy effects.

4 Experiment 2

To clarify the results of Experiment 1, we designed a second human subject experiment [62], which slightly modified the design of our initial experiment. First, while in Experiment 1 complex references all followed a uniform pattern (“Look at that {color} {shape}”), in this experiment we deviated from that pattern when there were multiple objects of the same color and shape, “Look at the {color} {shape} on your {direction relative to the person}” (e.g. “Look at the red tower on your right”).

Second, to encourage faster reaction times, we implemented a reaction time based point system. At the top of each survey page, participants were shown the number of points gained in the previous trial. For each video the participant would receive $15 - t$ points if correct, where t is the time in seconds taken to click on the object from when the video began. All videos were six seconds in length, including padding before and after the robot’s communicative act.

In this experiment we again examined four core hypotheses:

- H1.** Counter to Experiment 1, in this experiment we hypothesized that participants would have equal accuracy regardless of what communication style was used, as all communication styles that were used allowed for full disambiguation.
- H2–4.** We hypothesized the same expected effects of reaction time and perceived effectiveness and likability that we hypothesized in Experiment 1.

4.1 Measures and Participants

The same measures used in Experiment 1 were used in this Experiment. In this experiment, internal reliability scores for Effectiveness were still high (Effectiveness $\alpha = 0.975$, Likability $\alpha = 0.963$). 48 participants were recruited from Amazon Mechanical Turk (25 M, 17 F, 6 NA; ages 18 to 66, $M = 33.95$, $SD = 9.67$). None had participated in any previous studies from our laboratory.

4.2 Results

Accuracy: We hypothesized (**H1**) that participants would have equal accuracy regardless of what communication style was used, as all communication styles that were used allowed for full disambiguation. In fact, in refutation of H1, our results provided extreme evidence *in favor* of an effect of communication style (Bf 5.157e13), as seen in Fig. 3a.

Post-hoc analysis provided extreme evidence for differences in accuracy, between the complex reference condition ($M = 0.73$, $SD = 0.45$) and both the mixed reality condition ($M = 0.927$, $SD = 0.261$) (Bf 3.774e6) and the complex reference + mixed reality condition ($M = 0.938$, $SD = 0.242$) (Bf 3.160e7). This suggests that the use of complex reference by itself was significantly less effective than mixed reality deictic gesture. This effect was also seen in Experiment 1; the other effects seen in Experiment 1 did not appear in this experiment.

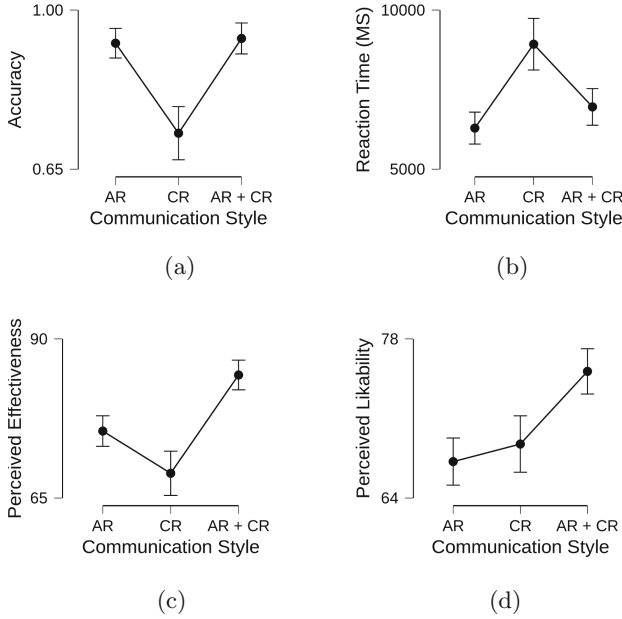


Fig. 3. (a) Effect of communication style (augmented reality (AR) vs complex reference (CR) vs both (AR+CR)) on participant accuracy. (b) Effect of communication style (augmented reality (AR) vs complex reference (CR) vs both (AR+CR)) on participant reaction time. (c) Effect of communication style (augmented reality (AR) vs complex reference (CR) vs both (AR+CR)) on perceived effectiveness (d) Effect of communication style (augmented reality (AR) vs complex reference (CR) vs both (AR+CR)) on perceived likability

Reaction Time: We hypothesized (**H2.1**) that reaction time would drop when mixed reality deictic gesture was used, as it would allow target referents to be disambiguated even before speech began, and (**H2.2**) that this difference in reaction time would be greater when a reference was ambiguous. While initial analysis provided strong evidence against an interaction between communication style and ambiguity (Bf 0.073, refuting (**H2.2**)), the evidence against a main effect of communication style was only anecdotal (Bf 0.665), prompting further exploration. Post-Hoc analysis provided moderate evidence in favor of a differences in reaction time between the complex reference condition ($M = 12.42$ s, $SD = 13.19$) and the mixed reality deictic gesture condition ($M = 9.69$, $SD = 10.78$) (Bf 4.204).

The extremely large standard deviations seen here led us to inspect our data, which showed a small number (about 5%) of our reaction time data points were very long, over 30s. Removing all reaction time datapoints for any participant with at least one outlier reaction time left 29 data points. Re-analyzing this subset of the data provided extreme evidence in favor of an effect of communication style (Bf 1.074e8), as seen in Fig. 3b. Post-hoc analysis provided extreme in favor of an effect of communication style specifically between the complex reference

condition ($M = 9.25$ s, $SD = 4.59$) and both the mixed reality deictic gesture condition ($M = 6.78$, $SD = 3.39$) (Bf 5.705e6) and the complex reference + mixed reality deictic gesture condition ($M = 7.25$, $SD = 3.40$) (Bf 1081.64). Figure 3b also appears to reflect a potential advantage of pure AR vs. AR paired with complex reference, but the post-hoc analysis provided anecdotal evidence against such an effect (Bf 0.705).

This suggests that the use of complex reference by itself may have taken longer to process than when augmented reality visualizations were used. This effect was not seen in Experiment 1, which failed to find evidence for or against the first hypothesis.

Effectiveness: We hypothesized (**H3.1**) that perceived effectiveness would be higher when mixed reality deictic gesture was used, especially (**H3.2**) when used in combination with complex reference, and (**H3.3**) when the target referent was ambiguous. Our results provided extreme evidence in favor of a main effect of communication style (Bf 3.42e14).

Post-hoc analysis provided extreme evidence in favor of a difference in perceived effectiveness between the mixed reality deictic gesture + complex reference condition ($M = 84.33$, $SD = 17.86$) and both the mixed reality deictic gesture condition ($M = 75.52$, $SD = 22.25$) (Bf 2.49e7) and the complex reference condition ($M = 68.89$, $SD = 22.98$) (Bf 1.65e10), as well as strong evidence in favor of a difference between the mixed reality deictic gesture and complex reference conditions (Bf 13.97). Specifically, our results show the same strong perceived ordering in effectiveness seen in Experiment 1: complex reference < mixed reality deictic gesture < complex reference + mixed reality deictic gesture, as seen in Fig. 3c. This confirms hypotheses **H3.1** and **H3.2**.

This specific ordering effect was also seen in Experiment 1; the other effects seen in Experiment 1 did not appear in this experiment.

Likability: We hypothesized that robots' perceived likability would correlate with their perceived effectiveness, and accordingly, that (**H4.1**) perceived likability would be higher when mixed reality deictic gesture was used, (**H4.2**) especially in combination with complex reference, and (**H4.3**) when the target referent was ambiguous. Our results provided extreme evidence in favor of a main effect of communication (Bf 1.64e6).

Post-hoc analysis provided extreme evidence in favor of a difference in likability between the mixed reality deictic gesture + complex reference condition ($M = 75.14$, $SD = 19.10$) and both the mixed reality deictic gesture condition ($M = 67.20$, $SD = 21.73$) (Bf 1.64e6) and the complex reference condition ($M = 68.75$, $SD = 20.25$) (Bf 632.99), as shown in Fig. 3d. This suggests that participants much more strongly liked the robot when it used both communication styles in combination, confirming hypothesis **H4.1**. This effect was also seen in Experiment 1; the other effects seen in Experiment 1 did not appear in this experiment.

4.3 Discussion

Our results suggest that mixed reality deictic gestures may be an accurate, efficient, likable, and effective communication strategy for human-robot interaction, much the same as traditional physical deictic gestures. In this section, we will discuss these results in detail, and leverage them to produce design guidelines for enabling mixed reality deictic gestures.

Objective Effectiveness of Mixed Reality Deictic Gesture: Our first and second hypotheses considered the objective effectiveness of mixed reality deictic gestures. Specifically, we hypothesized that while we did not expect there to be significant advantages in accuracy (**H1**), we did expect (**H2.1**) that the speed at which participants would be able to identify the robot’s target referent would be better when mixed reality deictic gesture was used, as it would allow target referents to be disambiguated even before speech began, and (**H2.2**) that this advantage in reaction time would be greater when a reference was ambiguous.

With respect to accuracy, our results suggest that the use of AR significantly increased accuracy over the use of bare complex reference, and that when complex reference was used by itself, participants incurred significant penalty to accuracy, *even though complex references were uniquely disambiguating and explicitly framed from participants’ point of view*. This surprising result refutes (**H1**), painting an even stronger picture of the benefits of mixed reality gesture.

With respect to reaction time, our results suggest that the use of AR significantly decreased reaction time over the use of bare complex reference, regardless of whether or not the target referent was ambiguous. This result supports (**H2.1**) and refutes (**H2.2**), again strengthening the overall utility of mixed reality deictic gesture, and demonstrating that the modifications made in this experiment over our previous work were an effective means to assess reaction time. However, additional study will be needed on this point, for two reasons. First, we suspect that advantages in the case of ambiguous referents will emerge as the number of distractors increases. Second, the number of temporal outliers that needed to be removed serves as a strong motivator for the need for the replication of this experiment in a live laboratory environment with realistic AR hardware, where such outliers would not be likely.

Subjective Perceptions of Mixed Reality Deictic Gesture: As in Experiment 1, our final hypotheses considered the subjective perception of mixed reality deictic gestures, hypothesizing (**H3.1/4.1**) that participants would perceive the robot to be more effective and likable when mixed reality deictic gesture was used, especially in combination with complex reference (**H3.2/4.2**), and for ambiguous referents (**H3.3/4.3**).

Our results suggest, as in Experiment 1, that the use of mixed reality deictic gesture improved perceived effectiveness *especially* when paired with complex reference (supporting (**H3.1**) and (**H3.2**) but refuting (**H3.3**)), and improved perceived likability *only* when paired with complex reference (supporting (**H4.2**) and partially supporting (**H4.1**) but refuting (**H4.3**)). These results emphasize that, like physical gesture, mixed reality deictic gesture should be used to

supplement rather than replace verbally expressive natural language. That being said, we would expect for very complex utterances that AR paired with referring expressions of reduced complexity would be preferred. Future work will be needed to determine if this is the case, and if so, how the tradeoff between referential complexity and positive perceptions of verbal expressivity should be quantified.

5 Current Implementation

In recent work, we have been working to enable generation of mixed reality deictic gestures on the Microsoft HoloLens. As described in recent work [60], we propose to decide when to generate mixed reality deictic gestures based on a variety of neurophysiological factors. This decision will be made by new components that will communicate with the Referring Expression Generation capabilities [77] of the Distributed, Integrated, Affect, Reflection, Cognition architecture [78, 79]. When it is decided to generate a gesture, this must be communicated to the HoloLens. We have established a bidirectional server-HoloLens connection using websockets, which we can use to send commands to show/hide visualizations over particular targets. When the HoloLens receives such a command, we render mixed reality deictic gestures through Unity over the appropriate location (Fig. 4). While we currently are mainly using ARTags to define object positions, in the future we hope to leverage object and pose recognition techniques to achieve these results without ARTags.



Fig. 4. HoloLens-projected arrow

6 Conclusion

We have explored the actual and perceived effectiveness of allocentric mixed reality deictic gestures in multi-modal robot communication. Our results suggest that these allocentric gestures may well be beneficial for human-robot communication in mixed reality environments, but highlight the importance of using them to complement complex referring expressions rather than purely as a replacement. Future work should seek to examine the tipping point at which referential overcomplexity overwhelms the subjective benefits of verbal expressivity. Second, it will be important to investigate a wider variety of mixed reality deictic gestures, with respect to both Sauppé and Mutlu [7] and our own [58] frameworks, and to investigate that wider array of gestures and evaluation criteria. We also hope to investigate the effect of different classes of mixed reality deictic gesture when used by robots of differing morphologies, e.g., robots that lack arms vs. robots that have arms they could use instead of (or in combination with)

allocentric gestures. Finally, after completing our integration on the Microsoft Hololens, we will attempt to replicate our experimental results on that system for increased external validity.

References

1. Ping, R., Goldin-Meadow, S.: Gesturing saves cognitive resources when talking about nonpresent objects. *Cogn. Sci.* **34**(4), 602–619 (2010)
2. Cook, S., Yip, T., Goldin-Meadow, S.: Gestures, but not meaningless movements, lighten working memory load when explaining math. *Lang. Cogn. Process.* **27**(4), 594–610 (2012)
3. Tfouni, L.V., Klatzky, R.L.: A discourse analysis of deixis: pragmatic, cognitive and semantic factors in the comprehension of ‘this’, ‘that’, ‘here’ and ‘there’. *J. Child Lang.* **10**(1), 123–133 (1983)
4. Marslen-Wilson, W., Levy, E., Tyler, L.K.: Producing interpretable discourse: the establishment and maintenance of reference. In: *Speech, Place, and Action* (1982)
5. Hanna, J.E., Tanenhaus, M.K.: Pragmatic effects on reference resolution in a collaborative task: evidence from eye movements. *Cogn. Sci.* **28**(1), 105–115 (2004)
6. Bangerter, A., Louwerse, M.M.: Focusing attention with deictic gestures and linguistic expressions. In: *Proceedings of the Annual Meeting of the Cognitive Science Society* (2005)
7. Sauppé, A., Mutlu, B.: Robot deictics: How gesture and context shape referential communication. In: *Proceedings of the International Conference on Human-Robot Interaction* (2014)
8. Levinson, S.C.: Deixis. In: *The Handbook of Pragmatics*, pp. 97–121. Blackwell (2004)
9. Norris, S.: Three hierarchical positions of deictic gesture in relation to spoken language: a multimodal interaction analysis. *Vis. Commun.* **10**(2), 1–19 (2011)
10. Bates, E.: *Language and Context: The Acquisition of Pragmatics*. Academic Press, Cambridge (1976)
11. Clark, E., Sengul, C.: Strategies in the acquisition of deixis. *J. Child Lang.* **5**(3), 457–475 (1978)
12. Iverson, J.M., Goldin-Meadow, S.: Gesture paves the way for language development. *Psychol. Sci.* **16**(5), 367–371 (2005)
13. Glenberg, A.M., McDaniel, M.A.: Mental models, pictures, and text: integration of spatial and verbal information. *Memory Cogn.* **20**(5), 458–460 (1992)
14. Gullberg, M.: Deictic gesture and strategy in second language narrative. In: *Workshop on the Integration of Gesture in Language and Speech* (1996)
15. De Angeli, A., Gerbino, W., Cassano, G., Petrelli, D.: Visual display, pointing, and natural language: the power of multimodal interaction. In: *AVI* (1998)
16. Goldin-Meadow, S.: The role of gesture in communication and thinking. *Trends Cogn. Sci. (TiCS)* **3**(11), 419–429 (1999)
17. Jancovic, M., Devoe, S., Wiener, M.: Age-related changes in hand and arm movements as nonverbal communication: some conceptualizations and an empirical exploration. *Child Dev.* **46**(4), 922–928 (1975)
18. Harrison, S.: The creation and implementation of a gesture code for factory communication. In: *Proceedings of the International Conference on Gesture in Speech and Interaction* (2011)

19. Brooks, A.G., Breazeal, C.: Working with robots and objects: revisiting deictic reference for achieving spatial common ground. In: Proceedings of the International Conference on HRI (2006)
20. Breazeal, C., Kidd, C., Thomaz, A.L., et al.: Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In: IROS (2005)
21. Salem, M., Kopp, S., Wachsmuth, I., Rohlfing, K., Joublin, F.: Generation and evaluation of communicative robot gesture. *Int. J. Soc. Robot.* **4**(2), 201–217 (2012)
22. Clair, A.S., Mead, R., Mataric, M.J.: Investigating the effects of visual saliency on deictic gesture production by a humanoid robot. In: RO-MAN (2011)
23. Okuno, Y., Kanda, T., Imai, M., Ishiguro, H., Hagita, N.: Providing route directions: design of robot's utterance, gesture, and timing. In: HRI (2009)
24. Liu, P., Glas, D., Kanda, T., Ishiguro, H., Hagita, N.: It's not polite to point: generating socially-appropriate deictic behaviors towards people. In: HRI (2013)
25. Cha, E., Kim, Y., Fong, T., Mataric, M.J.: A survey of nonverbal signaling methods for non-humanoid robots. *Found. Trends Robot.* **6**(4), 211–323 (2018)
26. Admoni, H., Weng, T., Scassellati, B.: Modeling communicative behaviors for object references in human-robot interaction. In: ICRA (2016)
27. Admoni, H., Weng, T., Hayes, B., Scassellati, B.: Robot nonverbal behavior improves task performance in difficult collaborations. In: HRI (2016)
28. Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., Joublin, F.: Effects of gesture on the perception of psychological anthropomorphism: a case study with a humanoid robot. In: Mutlu, B., Bartneck, C., Ham, J., Evers, V., Kanda, T. (eds.) *ICSR 2011. LNCS (LNAI)*, vol. 7072, pp. 31–41. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-25504-5_4
29. Salem, M., Kopp, S., et al.: Towards an integrated model of speech and gesture production for multi-modal robot behavior. In: RO-MAN (2010)
30. Holladay, R.M., Srinivasa, S.S.: RoGuE: robot gesture engine. In: Proceedings of the AAAI Spring Symposium Series (2016)
31. Whitney, D., Rosen, E., MacGlashan, J., Wong, L.L., Tellex, S.: Reducing errors in object-fetching interactions through social feedback. In: ICRA (2017)
32. Fang, R., Doering, M., Chai, J.Y.: Embodied collaborative referring expression generation in situated human-robot interaction. In: HRI (2015)
33. Piwek, P.: Saliency in the generation of multimodal referring acts. In: Proceedings of the International Conference on Multimodal Interfaces, pp. 207–210. ACM (2009)
34. Gatt, A., Paggio, P.: Learning when to point: a data-driven approach. In: Proceedings of International Conference on Computational Linguistics (COLING), pp. 2007–2017 (2014)
35. Van Der Sluis, I.F.: Multimodal reference, studies in automatic generation of multimodal referring expressions. Ph.D. thesis, University of Tilburg (2005)
36. Clark, H.: Coordinating with each other in a material world. *Discourse Stud.* **7**(4), 507–525 (2005)
37. Zhou, F., Duh, H.B.L., Billingham, M.: Trends in augmented reality tracking, interaction and display: a review of ten years of ISMAR. In: ISMAR (2008)
38. Van Krevelen, D., Poelman, R.: A survey of augmented reality technologies, applications and limitations. *Int. J. Virtual Reality* **9**(2), 1–20 (2010)
39. Billingham, M., Clark, A., Lee, G.: A survey of augmented reality. *Found. Trends Hum.-Comput. Interact.* **8**(2–3), 73–272 (2015)
40. Milgram, P., Zhai, S., Drascic, D., Grodski, J.: Applications of augmented reality for human-robot communication. In: Proceedings of IROS (1993)

41. Green, S., Billinghamurst, M., Chen, X., et al.: Human-robot collaboration: a literature review and augmented reality approach in design. *Int. J. Adv. Robot. Syst.* **5**(1), 1–18 (2008)
42. Williams, T., Szafor, D., Chakraborti, T.: The reality-virtuality interaction cube. In: *VAM-HRI* (2019)
43. Williams, T., Szafor, D., Chakraborti, T., Ben Amor, H.: Virtual, augmented, and mixed reality for human-robot interaction. In: *Companion of the HRI* (2018)
44. Williams, T., Szafor, D., Chakraborti, T., Amor, H.B.: Report on the 1st international workshop on virtual, augmented, and mixed reality for human-robot interaction (VAM-HRI). *AI Mag.* **39**(4), 64 (2018)
45. Peters, C., Yang, F., Saikia, H., Li, C., Skantze, G.: Towards the use of mixed reality for HRI design via virtual robots. In: *VAM-HRI* (2018)
46. Schönheits, M., Krebs, F.: Embedding AR in industrial HRI applications. In: *VAM-HRI* (2018)
47. Sportillo, D., Paljic, A., Ojeda, L., Partipilo, G., Fuchs, P., Roussarie, V.: Training semi-autonomous vehicle drivers with extended reality. In: *VAM-HRI* (2018)
48. Hedayati, H., Walker, M., Szafor, D.: Improving collocated robot teleoperation with augmented reality. In: *International Conference on HRI*, pp. 78–86. ACM (2018)
49. Ganesan, R.K., Rathore, Y.K., Ross, H.M., Amor, H.B.: Better teaming through visual cues. *IEEE Robot. Autom. Mag.* **25**(2), 59–71 (2018)
50. Chakraborti, T., Sreedharan, S., Kulkarni, A., Kambhampati, S.: Alternative modes of interaction in proximal human-in-the-loop operation of robots. arXiv preprint [arXiv:1703.08930](https://arxiv.org/abs/1703.08930) (2017)
51. Zu Borgesen, S., Renner, P., Lier, F., et al.: Improving human-robot handover research by mixed reality techniques. In: *VAM-HRI* (2018)
52. Walker, M., Hedayati, H., Lee, J., Szafor, D.: Communicating robot motion intent with augmented reality. In: *Proceedings of HRI*, pp. 316–324. ACM (2018)
53. Rosen, E., et al.: Communicating robot arm motion intent through mixed reality head-mounted displays. arXiv preprint [arXiv:1708.03655](https://arxiv.org/abs/1708.03655) (2017)
54. Amor, H.B., Ganesan, R.K., Rathore, Y., Ross, H.: Intention projection for human-robot collaboration with mixed reality cues. In: *VAM-HRI* (2018)
55. Sibirtseva, E., et al.: A comparison of visualisation methods for disambiguating verbal requests in human-robot interaction. In: *Proceedings of RO-MAN* (2018)
56. Perlmutter, L., Kernfeld, E., Cakmak, M.: Situated language understanding with human-like and visualization-based transparency. In: *RSS* (2016)
57. Reardon, C., Lee, K., Fink, J.: Come see this! augmented reality to enable human-robot cooperative search. In: *International Symposium on Safety, Security, and Rescue Robotics* (2018)
58. Williams, T., Tran, N., Rands, J., Dantam, N.T.: Augmented, mixed, and virtual reality enabling of robot deixis. In: Chen, J.Y.C., Fragomeni, G. (eds.) *VAMR 2018*. LNCS, vol. 10909, pp. 257–275. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91581-4_19
59. Williams, T.: A framework for robot-generated mixed-reality deixis. In: *VAM-HRI* (2018)
60. Hirshfield, L., Williams, T., Sommer, N., Grant, T., Gursoy, S.V.: Workload-driven modulation of mixed-reality robot-human communication. In: *ICMI Workshop on Modeling Cognitive Processes from Multimodal Data* (2018)
61. Williams, T., Bussing, M., Cabrol, S., Boyle, E., Tran, N.: Mixed reality deictic gesture for multi-modal robot communication. In: *HRI* (2019)
62. Williams, T., Bussing, M., Cabrol, S., Lau, I.: Toward allocentric mixed-reality deictic gesture. In: *VAM-HRI* (2019)

63. Gureckis, T.M., et al.: psiTurk: an open-source framework for conducting replicable behavioral experiments online. *Behav. Res. Methods* **48**(3), 829–842 (2016)
64. Huang, C.M., Mutlu, B.: Modeling and evaluating narrative gestures for humanlike robots. In: *Robotics: Science and Systems (RSS)*, pp. 57–64 (2013)
65. Bartneck, C., Kulić, D., Croft, E., Zoghbi, S.: Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Robot. (IJSR)* **1**(1), 71–81 (2009)
66. JASP Team: JASP (version 0.8.5.1) [computer software] (2018)
67. Wagenmakers, E., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J.: Bayesian inference for psychology, part II. *Psychon. Bull. Rev.* **25**(1), 58–76 (2018)
68. Crowder, M.J.: *Analysis of Repeated Measures*. Routledge, New York (2017)
69. Morey, R., Rouder, J.: Bayesfactor (version 0.9. 9) (2014)
70. Rouder, J.N., Morey, R.D., Speckman, P.L., Province, J.M.: Default Bayes factors for ANOVA designs. *J. Math. Psychol.* **56**(5), 356–374 (2012)
71. Mathôt, S.: Bayes like a Baws: interpreting Bayesian repeated measures in JASP [blog post], May 2017
72. Jeffreys, H.: Significance tests when several degrees of freedom arise simultaneously. *Proc. Roy. Soc. London Ser. A Math. Phys. Sci.* **165**(921), 161–198 (1938)
73. Westfall, P.H., Johnson, W.O., Utts, J.M.: A Bayesian perspective on the bonferroni adjustment. *Biometrika* **84**(2), 419–427 (1997)
74. Berger, J.O., Pericchi, L.R.: The intrinsic Bayes factor for model selection and prediction. *J. Am. Stat. Assoc.* **91**(433), 109–122 (1996)
75. Huettig, F., Rommers, J., Meyer, A.S.: Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychol.* **137**(2), 151–171 (2011)
76. Moratz, R., Ragni, M.: Qualitative spatial reasoning about relative point position. *J. Vis. Lang. Comput.* **19**(1), 75–98 (2008)
77. Williams, T., Scheutz, M.: Referring expression generation under uncertainty: algorithm and evaluation framework. In: *INLG* (2017)
78. Scheutz, M., Williams, T., Krause, E., Oosterveld, B., Sarathy, V., Frasca, T.: An overview of the distributed integrated cognition affect and reflection DIARC architecture. In: Aldinhas Ferreira, M.I., Silva Sequeira, J., Ventura, R. (eds.) *Cognitive Architectures. ISCASE*, vol. 94, pp. 165–193. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-97550-4_11
79. Scheutz, M., Briggs, G., Cantrell, R., Krause, E., Williams, T., Veale, R.: Novel mechanisms for natural human-robot interactions in the DIARC architecture. In: *Proceedings of the 2013 AAAI Workshop on Intelligent Robot Systems* (2013)