Toward Allocentric Mixed-Reality Deictic Gesture

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

Matthew Bussing
MIRRORLab
Colorado School of Mines
Golden, CO, USA
mbussing@mines.edu

Sebastian Cabrol
MIRRORLab
Colorado School of Mines
Golden, CO, USA
cabrol@mines.edu

Ian Lau
MIRRORLab
Colorado School of Mines
Golden, CO, USA
lau@mines.edu

***Abstract*—Research has shown that robots that use physical deictic gestures such as pointing enable more effective and natural interaction. However, it is not yet clear whether these benefits hold true for new forms of deictic gesture that become available in mixed-reality environments. In previous work, we presented a human-subject study suggesting that these benefits may indeed translate in the case of allocentric mixed-reality gestures, in which target referents are picked out in users' fields of view using annotations such as circles and arrows, especially when those gestures are paired with complex referring expressions. In this paper we provide additional evidence for this hypothesis through a second experiment that addresses potential confounds from our original experiment.**

***Index Terms*—Mixed Reality, Augmented Reality, Deixis, Natural Language Generation, Human-Robot Interaction**

I. Introduction

Robots are already being deployed in factories, hospitals, and search-and-rescue operations. Robots in these domains need to be able to communicate with their human users and teammates in a way that is effective and natural, while minimizing the need for special training. Accordingly, HRI researchers have sought to allow people to communicate with robots the same way they do with other people, through natural language and gesture. These two modes of communication are typically paired together, as gesture facilitates fluent speech, communication of abstract concepts, and deictic reference to nearby objects. Deictic gestures in particular are used by humans use to draw attention to the region of the environment containing their target referent.

HRI researchers have studied how robots might generate types of deictic gestures beyond pointing, including presenting, exhibiting, touching, grouping, and sweeping, and how such robot-generated gestures are perceived by humans [74]. As robots' capabilities increase, it will be just as important to study any new forms of gesture available to robot (and how those new forms are perceived). One way in which robots' capabilities are currently increasing is through their integration with augmented and mixed reality technologies [93], [94], which serve to increase not only the flexibility of users' control over robots, but also the expressivity of users' view into those robots' internal states [92], through visualizations rendered onto users' Augmented Reality Head-Mounted Displays that reflect information from the robot's internal state.

As a simple example, consider the case of a UAV communicating with human teammates about the location of a disaster victim. Purely using natural language, a UAV might use an utterance such as "There is an injured person behind the fourth tree to the right of the tall blue pylon." Such an utterance is complex, verbose, may require significant spatial reasoning capabilities to produce and may require sustained attention to interpret. While typically robots would use physical deictic gestures to reduce this complexity, enabling the robot to generate an utterance like "There is an injured person behind [point] that tree", it is unlikely that a UAV in this scenario will have an arm mounted on it, and the robot may not be easily visible from its' teammates' positions. In this situation, if the UAV's teammate are wearing AR-HMDs, the UAV might instead be able to simply draw a circle around the relevant tree, stating "There is an injured person behind [circle] that tree". This type of *Mixed Reality Deictic Gesture* would leverage the UAV's ability to manipulate its teammate's Mixed Reality Environment to achieve the same communicative goals as a physical gesture would have.

In recent work, we presented the first conceptual framework for categorizing the different types of Mixed Reality Deictic Gestures that may be used in Mixed Reality Human-Robot Interaction, as well as the dimensions along which such categories of gestures are expected to differ [89], [95], and presented the first systematic empirical examination [90] of the effectiveness and perception of one category of such gestures, *allocentric* gestures (in which circles, arrows, or other annotations are drawn into teammates' fields of view without connection to the robot generator). However, as we will describe, elements of the experimental design presented in that paper hindered the interpretation of their experimental results. In this paper, we present a new experiment that addresses those confounding factors to produce more straightforward results. As we will describe, our results suggest that allocentric mixed-reality deictic gestures are more accurate and efficient than bare complex reference, and *when paired with complex reference* improve perceived effectiveness and likability.

In Section II, we provide a brief survey of previous work exploring 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 Section III, we then describe the design of a human subject experiment 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 such gesture; a study designed to serve as a bridge towards future studies with real AR hardware. We the present the results of that experiment in Section IV. Finally, in Section V we discuss the implications of our experiment and suggest possible design guidelines for robot designers before concluding in Section VI.

II. Related Work

A. Human Deictic Gesture

Deixis is one of the most crucial pieces of human-human communications [51], [59], as well as one of the oldest, both anthropologically and developmentally. Unlike many other aspects of human communication, there are clear analogues of deictic gesture in the animal kingdom (e.g., the signaling capabilities of animals in the presence of predators) [54], [62]. However, deixis itself appears to be a predominantly human capability. While there has been some recent evidence suggesting that our closest relatives, primates, are able to point [40], [43], [49], [63], [75] and use other communicative gestures [66], evidence of pointing among other species is a relatively recent development in the literature [48]. This divergence in capability may exist in part because deixis may require relatively sophisticated capabilities involving modeling of attentional states and theory of mind [14], [26], [61]. Not only does reasoning about the feasibility and effectiveness of deictic gesture require about perspective taking, but more fundamentally, deictic gesture serves to direct an interlocutor's attention from where it is to where it should be; recognizing that an interlocutor's attention is not where you desire it to be is a complex capability indeed[1].

In contrast, humans point while speaking even from infancy, with deictic gesture beginning around 9-12 months [8], and general deictic reference mastered around age 4 [21]. Deictic gestures have been shown to be a powerful technique for language learners, as they allow speakers to communicate their intended referents before being able to do so in language, in the same way that 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 has been demonstrated to be a strong predictor of changes in language development [45]. 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 [25], [32], [33], [35], [46], and as deictic gesture allows for communication in environments in which verbal communication would be difficult or impossible, such as in noisy factory environments [37]. Accordingly, it is no surprise that Human-Robot Interaction researchers have sought to enable this effective and natural communication strategy in robots.

[1]We acknowledge that this is one perspective; many animal psychology researchers instead view pointing as a cognitively simple phenomenon explainable in operant terms [50], [53]. We direct the reader to the work of Krause et al. [48], which presents a comprehensive picture of this debate, which is beyond the scope of this paper.

B. Robot Deictic Gesture

Within the human-robot interaction literature, there has been widespread evidence for the effectiveness of robots' use of physical deictic gesture[2]. Specifically, 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 [15], and that robots' use of deictic gesture improves both subsequent human recall and human-robot rapport [13]. This effectiveness has been demonstrated across different contextual scales as well, including gestures to nearby objects on a tabletop [73], gestures to larger regions of space between the robot and its interlocutor [20], and gesture to large-scale spatial locations during direction-giving [60]. Furthermore, this effectiveness has shown to be especially true when gestures are generated in socially appropriate ways [52]. Research has also shown that robots' use of deictic gesture is especially effective when paired with *deictic gaze*, in which a robot (actually or ostensibly) shifts its gaze towards its intended referent [1], [2], [20], and that this is especially effective when gaze and gesture are appropriately coordinated [71]. Also of interest is a recent survey from Cha et al., in which deictic gaze and gesture are discussed within the context of a wide variety of nonverbal signaling mechanisms [16]. These findings have motivated a variety of technical approaches to deictic gesture generation [41], [42], [72], [88], as well as a number of approaches for integrating gesture generation with natural language generation [27] (see also [30], [31], [67], [83]).

Of particular interest to us is the work of Sauppé and Mutlu [74]. Building off the work of Clark, who showed that humans use many deictic gestures beyond pointing [22], 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 the design used by Sauppé and Mutlu.

C. Augmented Reality for HRI

Although research on augmented and mixed reality have been steadily progressing over the past several decades [5], [6], [12], [84], [96], there has been relatively little work using augmented reality (AR) technologies to facilitate human-robot interactions (despite a number of papers over the past twenty-five years highlighting the advantages of doing so [34], [57]).

[2]While there has also been significant work on robot *understanding* of human deictic gesture [56], we focus on robots' *generation* of such gestures.

Recently, however, research at the intersection of these fields has begun to dramatically increase [93], [94]. Recent work in this area includes approaches using AR for robot design [65], calibration [76], and training [79], and for communicating robots' perspectives [38], intentions [4], [17]–[19], [29] and trajectories [28], [69], [86], [97].

Most relevant to this paper are recent works on aligning human and robot perspective to enable more effective robot communication. Amor et al., for example, demonstrate the use of a projector to project instructions and highlight task-relevant objects within a constrained and highly structured task environment shared by robot and human teammates. In that work, however, no natural language generation is used, and projected visualizations are cast as part of the task environment, rather than as part of the robot's communication [3] (see also [4], [29]). Even more closely related, Sibirtseva et al. present an approach in which, as a human teammate describes a target referent to a robot, the robot's maintained distribution over possible intended referents is visualized by circling remaining reference candidates in the user's AR HMD [77] (see also similar work in VR from Perlmutter et al. [64]). This is closer to our area of interest, as the visualizations used in this work are explicitly used to pick out referential candidates, and are explicitly cast as being from the robot's perspective. However, we note that this is *passive* communication, as the robot is generating a backchannel response to the human's communication, whereas we are interested in robots' use of AR as a channel for *active* communication regarding its own intended referents. Moreover, Sibirtseva et al. were principally concerned with the tradeoffs between tablet, projector, and HMD-based AR visualizations, rather than on the impact of contextual factors. Also of interest is recent work from Reardon et al., in which a robot draws the trajectory a human teammate should take onto their field of view, and highlights the intended targets of that trajectory [68]. This work takes a more active communication approach than the work of Sibirtseva et al., but like Sibirtseva et al., Reardon et al. operate outside the context of language-based robot communication.

Finally, this work builds directly off of our own previous work [89], [95] (see also [39]), in which we presented a conceptual framework for categorizing the space of deictic gestures available in Mixed-Reality human-robot interactions, including both traditional physical gestures and purely 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. We then present an initial analysis hypothesizing how these combinations of potential gestures would differ along eleven dimensions, including privacy, cost, and legibility. This framework is especially valuable for our research as, in conjunction with the work of Sauppé and Mutlu [74], 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 the next section, we will present a set of such hypotheses, and a human subject experiment designed to investigate them. This experiment build off our previous experimental work on this topic [90], while addressing some methodological shortcomings of that previous work. While in this paper we will only examine allocentric gestures, we have designed our experiment so as to allow all of the gestural categories in our conceptual framework to be examined in future experiments using the same paradigm.

III. Experiment

To better understand the impact of mixed reality deictic gesture as a new modality for robot communication, and its interaction with natural language, we designed a human subject experiment in which participants viewed a robot referring to objects within a visual scene using natural language, mixed reality deictic gesture, or both modalities in combination. In previous work [90], we presented an initial experiment on this topic, modeled on the seminal evaluation of physical robot gesture presented by Sauppé and Mutlu [74]. In that work, all complex referring expressions generated by the robot followed a common template. This allowed for tight comparison with certain aspects of Sauppé and Mutlu's work, and for consistency of utterance lengths, but caused some utterances to be sufficiently ambiguous to be not uniquely resolveable. In this work, in order to remove these complications, we eliminated this constraint on utterance length. All aspects of our experimental design received IRB approval.

A. Experimental Design

Following Sauppé and Mutlu, we used a within-subject design, in which participants watched a robot refer to a series of twelve objects using different communication strategies.

1) 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") or, when there were multiple objects of the same color and shape (cf. our previous experimental design [90]), "Look at the {color} {shape} on your {direction relative to the person}" (e.g. "Look at the red tower on your right"). 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"; a pattern similar to the gestural conditions used by Sauppé and Mutlu. For the final third of the objects, the robot used both complex reference *and* mixed reality deictic gesture, circling the target and then generating a complex reference as described above; a pattern similar to a combination of the gestural and fully articulated conditions used by Sauppé and Mutlu.

Fig. 1: Task Environment, with simulated AR visualization

2) 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. 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 distance[3], yielding a total of twelve (3x2x2) experimental conditions.

B. Procedure

Participants were recruited online using Amazon's Mechanical Turk platform, and directed towards a psiTurk experimental environment [36][4]. After providing informed consent and providing demographic information[5], 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 660ms before speech onset, based on the gestural timing model presented by Huang and Mutlu [44] and leveraged by Sauppé and Mutlu [74]. 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. 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. This reward regimen was implemented in order to encourage participants to respond as quickly as possible, in order to study reaction times.

C. Hypotheses

We examined four core hypotheses:

- H1 We hypothesized that participants would have have equal accuracy regardless of what communication style was used, as all communication styles that were used allowed for full disambiguation.
- 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 advantage in reaction time would be greater 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 extent to which participants liked the robot 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 conjunction with complex reference, and **(H4.3)** for ambiguous targets.

D. Measures

To assess these hypotheses, objective and subjective measures were used. All measures were collected once per video.

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

2) 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).

3) Effectiveness: A subjective measure of robot *effectiveness* was gathered using a modified version of the Gesture Perception scale presented by Sauppé and Mutlu [74]. Our

[3] We did not expect to see any effects of distance, but decided to include 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.

[4] Mechanical Turk is more successful than traditional university studies at broad demographic sampling [24], though it still has population biases [81].

[5] In online experiments, it is valuable to collect demographic data pre-task to prevent participants who do not meet age requirements from participating.

modified version asked participants to evaluate each of the following statements by clicking a point anywhere along a seven-point Likert-type scale:

1) The robot used its speech and/or mixed reality deictic gesture effectively.
2) The robot's speech and/or mixed reality deictic gesture helped me to identify the object.
3) The robot's speech and/or mixed reality deictic gesture was appropriate for the context.
4) The robot's speech and/or mixed reality deictic gesture was easy to understand.

Each participants' scores for a single 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 α = 0.975.

4) Likability: A subjective measure of robot *likability* was gathered using the Godspeed II Likability scale [7]. 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 a single video were transformed to a range of 0-100 and averaged. A reliability analysis indicated very high internal reliability (Cronbach's α = 0.963).

E. Participants

48 participants were recruited from Amazon Mechanical Turk (25 M, 17 F, 6 NA). Participants ranged in age from 18 to 66 (M=33.95,SD=9.67). None had participated in any previous studies from our laboratory.

F. Analysis

Data analysis was performed within a Bayesian analysis framework using the JASP 0.9.2 [82] software package, using the default settings as justified by Wagenmakers et al. [85]. All data files are available at [Tobeuploadeduponacceptance]. For each measure, a repeated measures analysis of variance (RM-ANOVA) [23], [58], [70] was performed, using communication style, ambiguity, and distance as random factors[6]. These analysis used the default prior settings made by JASP, in order to meet formal desiderata [9]. Baws factors [55] 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, i.e.

$$\frac{\sum_{m \in M | e \in m} P(m|data)}{\sum_{m \in M | e \notin m} P(m|data)},$$

where e is an effect under consideration, and m is a candidate model in the space of candidate models M.

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 [47], [87] with a default Cauchy prior (center=0, r=$\frac{\sqrt{2}}{2}$=0.707).

While the Bayesian statistical approach has become increasingly common in the Cognitive Science and Psychology communities, it is still rare in the Human-Robot Interaction community, and as such we will briefly describe the benefits of this approach. First, the use of a Bayesian approach to statistical analysis provides some robustness to sample size (as it is not grounded in the central limit theorem). Second, the Bayesian approach allows for investigators to examine the evidence for and against hypotheses (whereas the frequentest approach can only quantify evidence towards rejection of the null hypothesis). Third, the Bayesian approach does not require reliance on p-values used in Null Hypothesis Significance Testing (NHST) which have recently come under considerable scrutiny [11], [78], [80]. Finally, the Bayesian framework facilitates the use of previous study results to construct informative priors so that our experiment may build upon our previous findings rather than starting anew.

IV. Results

A. Accuracy

We hypothesized **(H1)** that participants would have 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)[7], as seen in Fig. 2

Post-hoc analysis provided extreme evidence for differences in accuracy specifically between the complex reference condition (M=0.73,SD=0.45) and both the mixed reality deictic gesture condition (M=0.927,SD=0.261) (Bf 3.774e6) and the complex reference + mixed reality deictic gesture 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 our original experiment;other effects seen in our original experiment (a main effect of object ambiguity and a number of interaction effects) did not appear in this experiment.

B. 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 analysis provided moderate evidence in

[6]Note here that again, we did not expect to see effects of distance based on any of our hypotheses; but we*would* expect to see effects of distance in future experiments in which we plan to compare real and simulated robotic arms. We wanted this experiment to be directly comparable to those future experiments, so we selected a set of analyses that could be performed in both the current and future experiments.

[7]Bayes Factors above 100 indicate extreme evidence in favor of a hypothesis [10]. Here, for example, our Baws Factor Bf of 5.157e13 suggests that our data were 5.157e13 times more likely to be generated under models in which communication style is included than under those in which it is not.

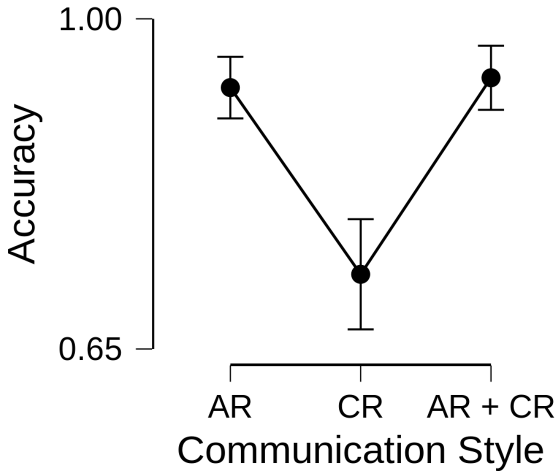

Fig. 2: Effect of communication style (Augmented Reality (AR) vs Complex Reference (CR) vs both (AR+CR)) on participant accuracy.

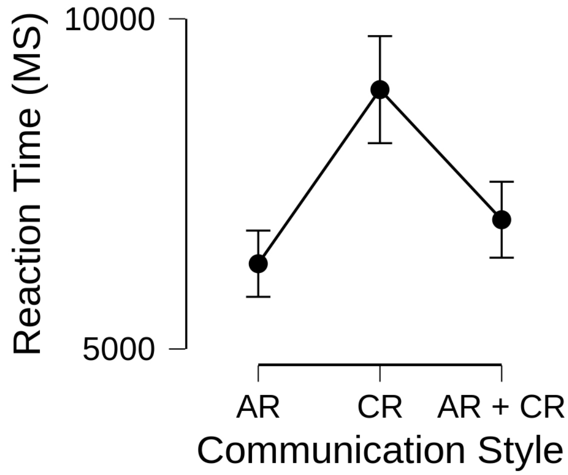

Fig. 3: Effect of communication style (Augmented Reality (AR) vs Complex Reference (CR) vs both (AR+CR)) on participant reaction time.

favor of a differences in reaction time between the complex reference condition (M=12.42 Seconds, 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 30 seconds. Removing all reaction time datapoints for any participant with at least one outlier reaction time left 29 data points. Reanalyzing this subset of the data provided extreme evidence in favor of an effect of communication style (Bf 1.074e8), as seen in Fig. 3. Post-hoc analysis provided extreme in favor of an effect of communication style specifically between the complex reference condition (M=9.25 Seconds,SD=4.59) and both the 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). Fig. 3 also appears to reflect a potential advantage of pure AR vs. AR paired with complex reference, but the post-hoc anlysis provided anecodtal 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 our original experiment, which failed to find evidence for or against the first hypothesis.

C. 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 our original: complex reference < mixed reality deictic gesture < complex reference + mixed reality deictic gesture, as seen in Fig. 4. This confirms hypotheses **H3.1** and **H3.2**.

This specific ordering effect was also seen in our original experiment; the other effects seen in our original experiment (a main effect of ambiguity and an interaction between communication style and ambiguity) did not appear in this experiment.

D. Likability

We hypothesized that robots' perceived likability would correlate with their effectiveness, and accordingly, that **(H4.1)** perceived likability would be higher when mixed reality deictic gesture was used, **(H4.2)** especially in conjunction 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

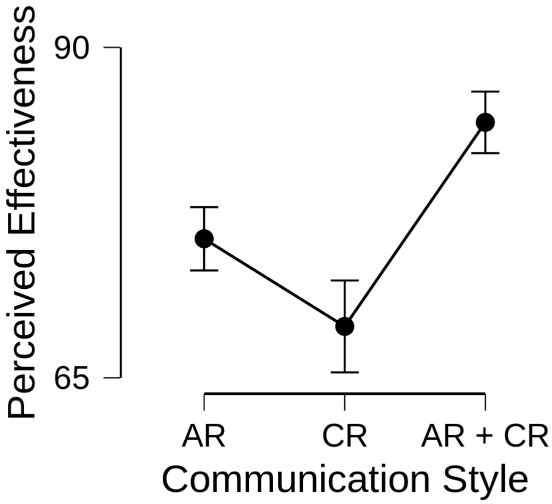

Fig. 4: Effect of communication style (Augmented Reality (AR) vs Complex Reference (CR) vs both (AR+CR)) on perceived effectiveness

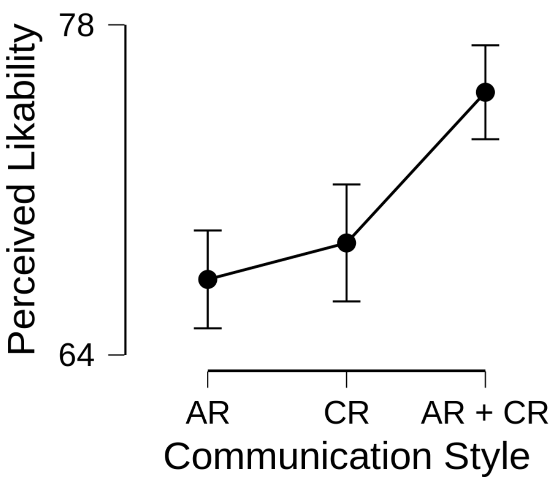

Fig. 5: Effect of communication style (Augmented Reality (AR) vs Complex Reference (CR) vs both (AR+CR)) on perceived likability

(M=68.75,SD=20.25) (Bf 632.99), as shown in Fig. 5. 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 our original experiment; the other effects seen in our original experiment (a main effect of ambiguity and an interaction between communication style and ambiguity) did not appear in this experiment.

V. 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.

A. *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 in fact, the use of AR significantly increased accuracy over the use of bare complex reference, and that when complex reference was used by itself, partipants 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**), instead painting an even stronger picture of the benefits of mixed reality deictic 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.

B. *Subjective Perceptions of Mixed Reality Deictic Gesture*

Our third and fourth hypotheses considered the subjective perception of mixed reality deictic gestures. Specifically, we hypothesized (**H3.1**) that participants would perceive the robot to be more effective when mixed reality deictic gesture was used, especially when used in conjunction with complex reference (**H3.2**), and when used to refer to inherently ambiguous referents (**H3.3**). In addition, we hypothesized (**H4.1**) that participants would perceive the robot to be more likable when mixed reality deictic gesture was used, especially when used in conjunction with complex reference (**H4.2**), and when used to refer to inherently ambiguous referents (**H4.3**)

Our results suggest, as in our original experiment, 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 serve to emphasize that, like physical gesture, mixed reality deictic gesture should be used to supplement rather than replace verbally expressive natural language (excepting extreme circumstances). 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.

VI. Conclusion

In this work we explored the actual and perceived effectiveness of allocentric mixed reality deictic gestures in multi-modal robot communication. Building off the findings presented in this paper, we see several promising directions for future work. First, 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 [74] and our own [95] frameworks, and to investigate that wider array of gestures with respect to the specific framework dimensions we previously highlighted. 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 conjunction with) allocentric gestures. Finally, we are currently in the process of implementing different mixed reality deictic gestures on the Microsoft Hololens. Once these gestures are integrated with our previous work on natural language generation [91], it will be critical to attempt to replicate the results of this experiment using that integrated system, for increased external validity.

References

[1] Henny Admoni, Thomas Weng, Bradley Hayes, and Brian Scassellati. Robot nonverbal behavior improves task performance in difficult collaborations. In *Proceedings of the 11th ACM/IEEE International Conference on Human Robot Interaction (HRI)*, pages 51–58. IEEE Press, 2016.

[2] Henny Admoni, Thomas Weng, and Brian Scassellati. Modeling communicative behaviors for object references in human-robot interaction. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, pages 3352–3359. IEEE, 2016.

[3] Heni Ben Amor, Ramsundar Kalpagam Ganesan, Yash Rathore, and Heather Ross. Intention projection for human-robot collaboration with mixed reality cues. In *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*, 2018.

[4] Rasmus S Andersen, Ole Madsen, Thomas B Moeslund, and Heni Ben Amor. Projecting robot intentions into human environments. In *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 294–301. IEEE, 2016.

[5] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent advances in augmented reality. Technical report, Naval Research Lab, Washington, DC, 2001.

[6] Ronald T Azuma. A survey of augmented reality. *Presence: Teleoperators & Virtual Environments*, 6(4):355–385, 1997.

[7] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics (IJSR)*, 1(1):71–81, 2009.

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

[9] M. J. Bayarri, J. O. Berger, A. Forte, and G. García-Donato. Criteria for Bayesian model choice with application to variable selection. *The Annals of Statistics*, 40:1550–1577, 2012.

[10] James O Berger and Luis R Pericchi. The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 91(433):109–122, 1996.

[11] James O Berger and Thomas Sellke. Testing a point null hypothesis: The irreconcilability of p values and evidence. *Journal of the American statistical Association*, 82(397):112–122, 1987.

[12] Mark Billinghurst, Adrian Clark, and Gun Lee. A survey of augmented reality. *Foundations and Trends in Human–Computer Interaction*, 8(2-3):73–272, 2015.

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

[14] Inge Bretherton, Sandra McNew, and Marjorie Beeghly-Smith. Early person knowledge as expressed in gestural and verbal communication: When do infants acquire a "theory of mind". *Infant social cognition*, pages 333–373, 1981.

[15] Andrew G Brooks and Cynthia Breazeal. 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 (HRI)*, pages 297–304. ACM, 2006.

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

[17] Tathagata Chakraborti, Andrew Dudley, and Subbarao Kambhampati. v2v communication for augmenting reality enabled smart huds to increase situational awareness of drivers. In *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*, 2018.

[18] Tathagata Chakraborti, Sarath Sreedharan, Anagha Kulkarni, and Subbarao Kambhampati. Alternative modes of interaction in proximal human-in-the-loop operation of robots. *arXiv preprint arXiv:1703.08930*, 2017.

[19] Mark Cheli, Jivko Sinapov, Ethan E. Danahy, and Chris Rogers. Towards an augmented reality framework for k-12 robotics education. In *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*, 2018.

[20] Aaron St Clair, Ross Mead, and Maja J Matarić. Investigating the effects of visual saliency on deictic gesture production by a humanoid robot. In *Proceedings of the 20th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 210–216. IEEE, 2011.

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

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

[23] Martin J Crowder. *Analysis of repeated measures*. Routledge, 2017.

[24] Matthew JC Crump, John V McDonnell, and Todd M Gureckis. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS one*, 8(3):e57410, 2013.

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

[26] Jill De Villiers. The interface of language and theory of mind. *Lingua*, 117(11):1858–1878, 2007.

[27] Rui Fang, Malcolm Doering, and Joyce Y Chai. Embodied collaborative referring expression generation in situated human-robot interaction. In

Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI), pages 271–278. ACM, 2015.

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

[29] Ramsundar Kalpagam Ganesan, Yash K Rathore, Heather M Ross, and Heni Ben Amor. Better teaming through visual cues. *IEEE Robotics & Automation Magazine*, 2018.

[30] Albert Gatt and Patrizia Paggio. What and where: An empirical investigation of pointing gestures and descriptions in multimodal referring actions. In *Proceedings of the 14th European Workshop on Natural Language Generation*, pages 82–91, 2013.

[31] Albert Gatt and Patrizia Paggio. Learning when to point: A data-driven approach. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING)*, pages 2007–2017, 2014.

[32] Arthur M Glenberg and Mark A McDaniel. Mental models, pictures, and text: Integration of spatial and verbal information. *Memory & Cognition*, 20(5):458–460, 1992.

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

[34] Scott A Green, Mark Billinghurst, XiaoQi Chen, and J Geoffrey Chase. Human-robot collaboration: A literature review and augmented reality approach in design. *International Journal of Advanced Robotic Systems*, 5(1):1, 2008.

[35] Marianne Gullberg. Deictic gesture and strategy in second language narrative. In *Workshop on the Integration of Gesture in Language and Speech*, pages 155–164. Applied Science and Engineering Laboratories, University of Delaware, 1996.

[36] Todd M Gureckis, Jay Martin, John McDonnell, Alexander S Rich, Doug Markant, Anna Coenen, David Halpern, Jessica B Hamrick, and Patricia Chan. psiturk: An open-source framework for conducting replicable behavioral experiments online. *Behavior research methods*, 48(3):829–842, 2016.

[37] Simon Harrison. The creation and implementation of a gesture code for factory communication. *Proceedings of the 2nd International Conference on Gesture in Speech and Interaction (GESPIN)*, 2011, 2011.

[38] Hooman Hedayati, Michael Walker, and Daniel Szafir. Improving collocated robot teleoperation with augmented reality. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 78–86. ACM, 2018.

[39] Leanne Hirshfield, Tom Williams, Natalie Sommer, Trevor Grant, and Senem Velipasalar Gursoy. Workload-driven modulation of mixed-reality robot-human communication. *Workshop on Modeling Cognitive Processes from Multimodal Data at the International Conference on Multimodal Interaction*, 2018.

[40] Catherine Hobaiter, David A Leavens, and Richard W Byrne. Deictic gesturing in wild chimpanzees (pan troglodytes)? some possible cases. *Journal of Comparative Psychology*, 128(1):82, 2014.

[41] Rachel M Holladay, Anca D Dragan, and Siddhartha S Srinivasa. Legible robot pointing. In *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 217–223. IEEE, 2014.

[42] Rachel M Holladay and Siddhartha S Srinivasa. Rogue: Robot gesture engine. In *Proceedings of the AAAI Spring Symposium Series*, 2016.

[43] William D Hopkins, Jared P Taglialatela, and David A Leavens. Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Animal behaviour*, 73(2):281–286, 2007.

[44] Chien-Ming Huang and Bilge Mutlu. Modeling and evaluating narrative gestures for humanlike robots. In *Robotics: Science and Systems (RSS)*, pages 57–64, 2013.

[45] Jana M Iverson and Susan Goldin-Meadow. Gesture paves the way for language development. *Psychological science*, 16(5):367–371, 2005.

[46] MerryAnn Jancovic, Shannon Devoe, and Morton Wiener. Age-related changes in hand and arm movements as nonverbal communication: Some conceptualizations and an empirical exploration. *Child Development*, pages 922–928, 1975.

[47] Harold Jeffreys. Significance tests when several degrees of freedom arise simultaneously. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, pages 161–198, 1938.

[48] Mark A Krause, Monique AR Udell, David A Leavens, and Lyra Skopos. Animal pointing: changing trends and findings from 30 years of research. *Journal of Comparative Psychology*, 2018.

[49] David A Leavens. Manual deixis in apes and humans. *Interaction Studies*, 5(3):387–408, 2004.

[50] David A Leavens. Joint attention: twelve myths. *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience*, pages 43–72, 2011.

[51] Stephen C Levinson. Deixis. In *The handbook of pragmatics*, pages 97–121. Blackwell, 2004.

[52] Phoebe Liu, Dylan F Glas, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. It's not polite to point: generating socially-appropriate deictic behaviors towards people. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 267–274. IEEE Press, 2013.

[53] Heidi Lyn, Patricia M Greenfield, Sue Savage-Rumbaugh, Kristen Gillespie-Lynch, and William D Hopkins. Nonhuman primates do declare! a comparison of declarative symbol and gesture use in two children, two bonobos, and a chimpanzee. *Language & communication*, 31(1):63–74, 2011.

[54] Peter Marler. Primate vocalization: affective or symbolic? In *Speaking of apes*, pages 221–229. Springer, 1980.

[55] S. Mathôt. Bayes like a baws: Interpreting bayesian repeated measures in JASP [blog post]. https://www.cogsci.nl/blog/interpreting-bayesian-repeated-measures-in-jasp, May 2017.

[56] Cynthia Matuszek, Liefeng Bo, Luke Zettlemoyer, and Dieter Fox. Learning from unscripted deictic gesture and language for human-robot interactions. In *AAAI*, pages 2556–2563, 2014.

[57] Paul Milgram, Shumin Zhai, David Drascic, and Julius Grodski. Applications of augmented reality for human-robot communication. In *Intelligent Robots and Systems' 93, IROS'93. Proceedings of the 1993 IEEE/RSJ International Conference on*, volume 3, pages 1467–1472. IEEE, 1993.

[58] RD Morey and JN Rouder. Bayesfactor (version 0.9. 9), 2014.

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

[60] Yusuke Okuno, Takayuki Kanda, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. Providing route directions: design of robot's utterance, gesture, and timing. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 53–60. ACM, 2009.

[61] Daniela K O'Neill. Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child development*, 67(2):659–677, 1996.

[62] Daniel Otte. Effects and functions in the evolution of signaling systems. *Annual Review of Ecology and Systematics*, 5(1):385–417, 1974.

[63] Marie Pelé, Valérie Dufour, Bernard Thierry, and Josep Call. Token transfers among great apes (gorilla gorilla, pongo pygmaeus, pan paniscus, and pan troglodytes): species differences, gestural requests, and reciprocal exchange. *Journal of Comparative Psychology*, 123(4):375, 2009.

[64] Leah Perlmutter, Eric Kernfeld, and Maya Cakmak. Situated language understanding with human-like and visualization-based transparency. In *Robotics: Science and Systems*, 2016.

[65] Christopher Peters, Fangkai Yang, Himangshu Saikia, Chengjie Li, and Gabriel Skantze. Towards the use of mixed reality for hri design via virtual robots. In *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*, 2018.

[66] Simone Pika, Katja Liebal, Josep Call, and Michael Tomasello. Gestural communication of apes. *Gesture*, 5(1):41–56, 2005.

[67] Paul Piwek. Salience in the generation of multimodal referring acts. In *Proceedings of the 2009 international conference on Multimodal interfaces*, pages 207–210. ACM, 2009.

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

[69] Eric Rosen, David Whitney, Elizabeth Phillips, Gary Chien, James Tompkin, George Konidaris, and Stefanie Tellex. Communicating robot arm motion intent through mixed reality head-mounted displays. *arXiv preprint arXiv:1708.03655*, 2017.

[70] Jeffrey N Rouder, Richard D Morey, Paul L Speckman, and Jordan M Province. Default bayes factors for anova designs. *Journal of Mathematical Psychology*, 56(5):356–374, 2012.

[71] Maha Salem, Friederike Eyssel, Katharina Rohlfing, Stefan Kopp, and Frank Joublin. Effects of gesture on the perception of psychological anthropomorphism: a case study with a humanoid robot. In *International Conference on Social Robotics*, pages 31–41. Springer, 2011.

[72] Maha Salem, Stefan Kopp, Ipke Wachsmuth, and Frank Joublin. Towards an integrated model of speech and gesture production for multi-modal robot behavior. In *19th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 614–619. IEEE, 2010.

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

[74] Allison Sauppé and Bilge Mutlu. Robot deictics: How gesture and context shape referential communication. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 342–349. ACM, 2014.

[75] E Sue Savage-Rumbaugh. Language as a cause-effect communication system. *Philosophical psychology*, 3(1):55–76, 1990.

[76] Manfred Schönheits and Florian Krebs. Embedding ar in industrial hri applications. In *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*, 2018.

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

[78] Joseph P Simmons, Leif D Nelson, and Uri Simonsohn. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 22(11):1359–1366, 2011.

[79] Daniele Sportillo, Alexis Paljic, Luciano Ojeda, Giacomo Partipilo, Philippe Fuchs, and Vincent Roussarie. Training semi-autonomous vehicle drivers with extended reality. In *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*, 2018.

[80] Jonathan AC Sterne and George Davey Smith. Sifting the evidence – what's wrong with significance tests? *Physical Therapy*, 81(8):1464–1469, 2001.

[81] Neil Stewart, Jesse Chandler, and Gabriele Paolacci. Crowdsourcing samples in cognitive science. *Trends in Cognitive Sciences (TiCS)*, 2017.

[82] JASP Team. Jasp (version 0.8.5.1)[computer software], 2018.

[83] Ielka Francisca Van Der Sluis. *Multimodal Reference, Studies in Automatic Generation of Multimodal Referring Expressions*. PhD thesis, University of Tilburg, 2005.

[84] DWF Van Krevelen and Ronald Poelman. A survey of augmented reality technologies, applications and limitations. *International journal of virtual reality*, 9(2):1, 2010.

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

[86] Michael Walker, Hooman Hedayati, Jennifer Lee, and Daniel Szafir. Communicating robot motion intent with augmented reality. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 316–324. ACM, 2018.

[87] Peter H Westfall, Wesley O Johnson, and Jessica M Utts. A bayesian perspective on the bonferroni adjustment. *Biometrika*, 84(2):419–427, 1997.

[88] David Whitney, Eric Rosen, James MacGlashan, Lawson LS Wong, and Stefanie Tellex. Reducing errors in object-fetching interactions through social feedback. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2017.

[89] Tom Williams. A framework for robot-generated mixed-reality deixis. In *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*, 2018.

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

[91] Tom Williams and Matthias Scheutz. Referring expression generation under uncertainty: Algorithm and evaluation framework. In *Proceedings of the 10th International Conference on Natural Language Generation (INLG)*, 2017.

[92] Tom Williams, Daniel Szafir, and Tathagata Chakraborti. The reality-virtuality interaction cube: A framework for conceptualizing mixed-reality interaction design elements for hri. In *Late Breaking Reports of the 14th ACM/IEEE International Conference on Human-Robot Interaction*, 2019.

[93] Tom Williams, Daniel Szafir, Tathagata Chakraborti, and Heni Ben Amor. Report on the 1st international workshop on virtual, augmented, and mixed reality for human-robot interaction (VAM-HRI). *AI Magazine*, 2018.

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

[95] Tom Williams, Nhan Tran, Josh Rands, and Neil T. Dantam. Augmented, mixed, and virtual reality enabling of robot deixis. In *Proceedings of the 10th International Conference on Virtual, Augmented, and Mixed Reality (VAMR)*, 2018.

[96] Feng Zhou, Henry Been-Lirn Duh, and Mark Billinghurst. Trends in augmented reality tracking, interaction and display: A review of ten years of ismar. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 193–202. IEEE Computer Society, 2008.

[97] Sebastian Meyer zu Borgsen, Patrick Renner, Florian Lier, Thies Pfeiffer, and Sven Wachsmuth. Improving human-robot handover research by mixed reality techniques. In *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*, 2018.