(Gestures Vaguely): The Effects of Robots' Use of Abstract Pointing Gestures in Large-Scale Environments

Annie Huang*
Colorado School of Mines
Golden, CO, USA
anniehuang@mines.edu

Alyson Ranucci*
Colorado School of Mines
Golden, CO, USA
aranucci@mines.edu

Adam Stogsdill*
RAIsonance
Greenwood Village, CO, USA
adam.stogsdill@gmail.com

Grace Clark
Colorado School of Mines
Golden, CO, USA
clarkgrace18@gmail.com

Keenan Schott
Colorado School of Mines
Golden, CO, USA
keenanschott@mines.edu

Mark Higger
Colorado School of Mines
Golden, CO, USA
mhigger@mines.edu

Zhao Han
University of South Florida
Tampa, FL, USA
zhaohan@usf.edu

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

ABSTRACT

As robots are deployed into large-scale human environments, they will need to engage in task-oriented dialogues about objects and locations beyond those that can currently be seen. In these contexts, speakers use a wide range of referring gestures beyond those used in the small-scale interaction contexts that HRI research typically investigates. In this work, we thus seek to understand how robots can better generate gestures to accompany their referring language in large-scale interaction contexts. In service of this goal, we present the results of two human-subject studies: (1) a human-human study exploring how human gestures change in large-scale interaction contexts, and to identify human-like gestures suitable to such contexts yet readily implemented on robot hardware; and (2) a human-robot study conducted in a tightly controlled Virtual Reality environment, to evaluate robots' use of those identified gestures. Our results show that robot use of Precise Deictic and Abstract Pointing gestures afford different types of benefits when used to refer to visible vs. non-visible referents, leading us to formulate three concrete design guidelines. These results highlight both the opportunities for robot use of more humanlike gestures in large-scale interaction contexts, as well as the need for future work exploring their use as part of multi-modal communication.

CCS CONCEPTS

• **Computer systems organization → Robotics**; • **Human-centered computing → Virtual reality**; **Empirical studies in interaction design**.

*The first three authors contributed equally to this paper.

HRI '24, March 11–14, 2024, Boulder, CO, USA

ACM ISBN 979-8-4007-0322-5/24/03.
https://doi.org/10.1145/3610977.3634924

KEYWORDS

Deixis, spatial reference, non-verbal communication, anthropomorphism, virtual reality (VR), human-robot interaction (HRI)

ACM Reference Format:
Annie Huang, Alyson Ranucci, Adam Stogsdill, Grace Clark, Keenan Schott, Mark Higger, Zhao Han, and Tom Williams. 2024. (Gestures Vaguely): The Effects of Robots' Use of Abstract Pointing Gestures in Large-Scale Environments. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24), March 11–14, 2024, Boulder, CO, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3610977.3634924

1 INTRODUCTION

Suppose that while talking to a colleague in your office, you were to mention that HRI 2024 was to be held in Boulder, Colorado. Even if you knew quite well the general direction that Boulder was from your current location, you would probably not turning to face Boulder, extend your arm precisely, and gaze intently at your office wall. Instead, you would likely wave your hand as if to say "elsewhere" – a purely abstract gesture not intended to be followed or used to establish joint attention. On the other hand, if you were to talk to your colleague about an object in front of you (perhaps a craft beer from Boulder, Colorado), you would likely not gesture vaguely, but instead use a deictic gesture like pointing or presenting, gaze at your referent directly, and perhaps even actively check to ensure your interlocutor was following your gaze and gesture, thus establishing shared attention.

In many cases, human selection of gestures to accompany referring expressions may be straightforward due simply to the limitations of human cognition. In many cases, unless we can see an object, or have a landmark such as mountains or a coastline to ground our sense of direction toward a far-off location, we have little to no idea of the heading along which target referents lie, and we would be unable to precisely gesture towards most referents without seconds or minutes of careful deliberate thought and geometric reasoning. As recent work on human-human gesture has demonstrated, this leads to a wide array of referring gestures being used

beyond precise deictic pointing, even in relatively small-scale interaction contexts with only modest environmental occlusion [27].

Robots, on the other hand, are not under the same limitations, and may in fact have precise metric knowledge of objects and locations that are not currently visible and potentially quite far away. In such cases, from the robot's perspective, a deictic gesture would be easy to generate, even though it would likely be hard to interpret and potentially confusing for human interlocutors. This disconnect is increasingly important as robots are moved from small-scale interaction contexts in which all candidate referents are readily visible, into large-scale interaction contexts where most referents are *not* currently visible, like hospitals, shopping malls, and other large-scale public environments.

As such, we argue that for robots to effectively and naturally use referring gestures in realistic, large-scale human environments, robots' nonverbal behaviors must be designed with sensitivity to what humans find to be natural, humanlike, and understandable *in those large-scale contexts*. In service of this goal, we make two key contributions in this work. First, we present the results of a human-human study conducted to understand how human gestures change as referring context expands, and to identify human-like gestures suitable to large-scale interaction contexts. Second, we present the results of a human-robot study conducted in a tightly controlled Virtual Reality environment, to evaluate robots' use of those identified gestures in large-scale interaction contexts. Our results from both ethics board approved studies show that robot use of Precise Deictic and Abstract Pointing gestures afford different types of benefits when used to refer to visible vs. non-visible referents, leading us to formulate three concrete design guidelines. These results highlight both the opportunities for robot use of more humanlike gestures in large-scale interaction contexts, and the need for future work exploring their use as part of multi-modal communication.

2 RELATED WORK

2.1 Human Gesture

Gestures are one of the most important channels used in human-human communication. They allow listeners to better understand the meaning and intentions behind speakers' utterances, both in typical dialogue and in contexts in which words cannot be used or in which interlocutors speak different languages [35, 53]. Gestures are especially useful in such contexts due to their visual nature; as Kendon [35] argues, gesture allows speech to convey additional *mental imagery* that persists even once the speaker has finished speaking. Moreover, the use of gestures plays a significant role in the *gesturer's* cognition [21], enabling speakers to work through and better articulate concepts, even if they are unable to see the gestures they are making [32]. Gestures are particularly common when a speaker is referencing spatial information [3], in part because of gestures' utility as a visuospatial information channel that can be used to supplement non-visuospatial speech [40].

As delineated by McNeill [40] human gestures can be divided into five main categories: deictics (which help pick out physical referents), iconics (which resemble physical shapes), metaphorics (which represent more abstract concepts), cohesives (abstract gestures used to metaphorically connect narrative elements), and beats (which do not reflect concepts, but instead provide emphasis and reflect tempo). While categories like *iconics*, which directly depict figural representations, most literally convey mental imagery, each of these categories conveys imagery or visuospatial information in some way. Deictic gestures like pointing, presenting, and sweeping, are particularly effective at conveying spatial information, and are often used during tasks with significant spatial components, such as giving directions [4] or describing room layouts [48].

But moreover, deictic reference, whether in the form of deictic language, gaze, or gesture, is a critical part of situated human-human communication [38, 41]. Deixis is one of the earliest forms of communication both anthropologically and developmentally. Beginning around 9-12 months, humans learn to point during speech [6], with mastery of deictic reference attained around age 4 [14]. Because deictic gestures allow speakers to pick out referents without using language (similar to how other gesture types allow communicators to express more abstract meanings not grounded to the environment), they are a robust technique for language learning. As a result, language development changes can be predicted through developmental changes in humans' deictic gestural skills [33]. Furthermore, humans continue to rely on the use of deictic gestures long past infancy as a major communicative skill due to its usefulness as a referential strategy in complex environments, such as noisy work environments [26], that require (or at least benefit from) more communication channels beyond speech [15, 19, 20, 22, 34].

Historically, deictic gesture has typically been studied in small-scale interaction contexts where humans must refer to visible objects, locations, and people. But as Enfield et al. [16] demonstrate in their study of referring language used in Laotian villages, a wider range of referring gesures can be observed if we consider larger interaction contexts. Enfield et al. [16], for example, highlight the use of "Big" points comprised of large full-arm gestures (used to point to specific locations in space) versus "Small" points with smaller movements and more complex hand movements (used to help resolve particularly ambiguous referents and refer to entities not currently visible). Recently, researchers have begun to more carefully analyze referring gestures used in other large-scale interaction contexts. For example, Higger et al. [27] recently presented a new taxonomy of referring gestures comprised of five distinct categories: three different types of deictic gestures used to achieve varying levels of disambiguation, a category of iconic gestures used for referring purposes, and a category of "Abstract Pointing" gestures comprised of *non-deictic* pointing gestures (e.g., pointing vaguely in some direction that may or may not actually lead towards the target referent). While this taxonomy captures the *types* of gestures used by humans to refer to non-visible objects, however, it does not explain the *criteria* that speakers use when deciding between gestures. Moreover, no work yet considers how these different forms of referring gestures might be deployed in *human-robot* interaction.

2.2 Robot Deictic Gesture

There is a long history of work on robot gesture generation in the Human-Robot Interaction community. In particular, due to the situated nature of human-robot communication, deictic gestures have been especially extensively studied in the HRI literature, including deictic gestures used to refer to objects in tabletop interactions [46, 47], deictic gestures used to refer to larger spatial

regions [13] and during direction-giving [42]. Deictics have been of particular interest in the context of situated, task-oriented human-robot communication. Robots' use of deictic gesture is effective at shifting attention in the same way as humans' use of deictic gesture [10], and robots' use of deictic gesture improves both subsequent human recall [31] and human-robot rapport [7]. Research has also shown that robots' use of deictic gesture is especially effective when paired with other nonverbal signaling mechanisms [12], such as *deictic gaze*, in which a robot (actually or ostensibly) shifts its gaze towards its intended referent [1, 2, 13], and that this is especially effective when gaze and gesture are appropriately coordinated [44]. All of these findings suggest that deictic gesture is a critical component across a wide breadth of pro-social HRI contexts, such as healthcare contexts (where researchers aim to reduce inequities in communities' health-related capabilities) and education contexts (where researchers aim to reduce inequities in communities' capabilities to sense, think, imagine, and play)[57].

Accordingly, these findings have motivated a variety of technical approaches for deictic gesture generation [29, 45, 52] and for integrating gesture generation with natural language generation [17, 18, 43, 50]). Recent work has even shown how robot gestures may be generated through interactive modalities like Augmented Reality [11, 23, 25, 49, 55] to unique effect. As in the human-human interaction literature, however, there has been little attention to referring gestures beyond precise deictic pointing.

2.3 Robot Abstract Gesture

Most research on abstract robot gestures focuses on beat gestures [9], iconic gestures [8, 30], and metaphoric gestures [30]. Many of these approaches have also looked at joint generation of deictic and abstract gestures [30, 31]. Yet these approaches have typically ignored the ways that abstract gestures might be used as part of referential communication in the way that deictic gestures are.

This may be due in part to the interaction contexts typically used in HRI research, in which a limited, finite, and visible set of objects are assumed to be under discussion, all of which can be assumed to be known to both human and robot, and which are typically located immediately in front of the robot or are at least visible in the environment (e.g., on a table [1, 2, 17, 28, 47] or screen [30], [cp. 42]). In such cases, the most natural gesture to accompany spatial language is a precise deictic pointing gesture, where the robot points and gazes directly at an object to allow interlocutors to achieve joint attention by following the robot's gesture and gaze.

In realistic task contexts, however, the space of possible objects is not limited to a finite set. As highlighted in work targeting linguistic reference understanding [54, 56], robots must also understand and generate references to objects and locations that are not currently visible (or in fact that may never have been seen or heard of before).

Based on the literature described above, there are at least two key research aims that will be critical for the HRI community to pursue as robots are deployed into larger-scale environments than those traditionally examined in laboratory-based HRI research. First, cognitive scientists must work to understand the factors that determine *when and why* humans typically use the different types of referring gestures delineated within Higger et al. [27]'s recent taxonomy. And second, roboticists must use those insights to design more humanlike gestures for use by robots in large-scale interaction contexts, and work to understand the objective performance and subjective perception of those gestures. As cognitive scientists and roboticists, we thus work to advance both research aims.

3 EXPERIMENT ONE

In our first study, we seek to answer our first key research question: **(RQ1)** When and why do humans use different types of referring gestures? To answer this question, we conducted an exploratory study following a within-subjects design.

3.1 Method

To investigate this research question, we designed a spatial reference task in which participants sequentially referred to a series of familiar objects and locations. The objects included in this list contained objects clearly visible in the experiment room, common landmarks within the building housing the experiment, other nearby buildings and landmarks, and commonly known US cities. The distribution of objects and locations in this set roughly followed a negative exponential curve, with many nearby referents included and few distant referents included.

Experimental participants engaged in this task in a dyadic context, in which the participant sat across from the experimenter, and was sequentially asked by the experimenter to describe the location of each object or location, with the experimenter only referring to each target by proper name (i.e., without themselves using any gaze or gestural behaviors or describing the target in any way). When referring to the objects, the participants were required to refer to them verbally but *were not required* to use gestures. If participants asked for clarification on how to describe an object, they were encouraged to describe the location of the object in the way that made the most sense to them.

Participants' gestural behaviors were videotaped using RGB and RGB-D cameras. All videos were coded by a primary rater, who categorized the gestures accompanying participants' spatial referring expressions as either more precise, more abstract, or absent. Here, deictic gestures such as pointing, sweeping, and presenting were categorized as more precise (cp. [47]), and all other gestures (including metaphoric, abstract, and beat gestures) were categorized as more abstract. When categorization was unclear, coding was determined through consultation with a secondary rater. Whenever a participant indicated that they were unfamiliar with one of the referents to be described, we removed their data for that referent. Moreover, two objects were removed completely from our analysis because the majority of participants were unfamiliar with their location. The remaining data is visualized in Fig. 1

Fourteen participants were recruited from a mid-sized US college campus to participate in this exploratory experiment. This produced a dataset of 254 recorded descriptions. Participants were paid $5 each for their participation. Examples of gestures coded as more precise vs. more abstract are shown in Fig. 2.

3.2 Analysis

After completing data collection, a Bayesian analysis was performed to understand the role of two key factors (referent visibility and

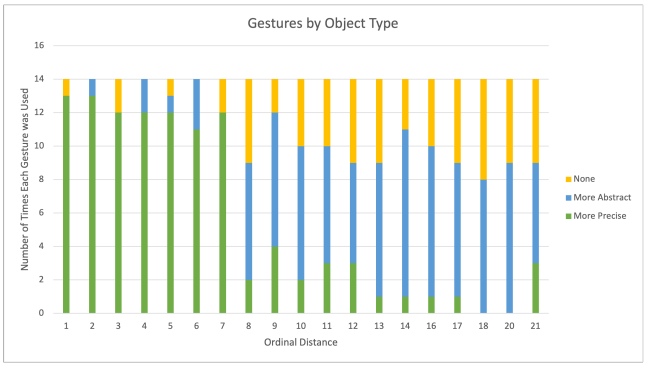

Figure 1: Gestures used in Experiment One. Target referents are ordered from left to right in increasing order of distance. A dramatic drop in the use of more precise gestures is observed for referents sufficiently far away to be no longer visible (i.e., those of rank eight and above), after which more abstract gestures are typically used, with a negative trend from that point onward in the use of more precise gestures.

Figure 2: Participant gestures coded as More Precise (left) and as More Abstract (right) in Experiment One.

referent distance) in predicting interactants' use of more abstract versus more precise referring gestures.

The Bayesian approach has several advantages [51] over the Frequentist approach although it is not yet as commonly used. Key advantages of this framework include (1) the ability to gather evidence in favor of the null hypothesis and, more generally, quantify the evidence for and against competing hypotheses; (2) the ability to engage in flexible sampling plans, e.g., to "peek" at data before sampling has concluded and use this to make decisions as to whether or not to continue collecting data.

We used the brms R package to fit and compare a series of General Linear Mixed Models, each with a different combination of *distance to target referent* (a log-scale continuous variable measured in feet), *target referent visibility* (a binary variable), and *speaker* (a categorical variable to account for individual differences) to predict gesture type (a categorical variable). All models used the logistic function as the model link function. After fitting these models, Bayes Inclusion Factors Across Matched Models were calculated [39] to quantify the relative evidence for inclusion versus noninclusion of each of these two factors and their potential interaction.

Here, Bayes Factors BF represent the ratio of evidence between the two competing hypotheses $\mathcal{H}_1$ and $\mathcal{H}_0$. For example, $BF_{10} = 5$ means that the data collected is 5 times more likely to occur under $\mathcal{H}_1$ than under $\mathcal{H}_0$. To interpret the results of our Bayes Factor analyses, we used the widely accepted interpretation scheme proposed by Lee and Wagenmakers [37]. Under this approach, evidence is considered anecdotal (inconclusive) for $BF \in [1/3, 3]$, moderate when $BF \in [3, 10]$ (or when $BF \in [1/3, 1/10]$), strong when $BF \in [10, 30]$ (or when $BF \in [1/10, 1/30]$), very strong when $BF \in [30, 100]$ (or when $BF \in [1/30, 1/100]$), and extreme when $BF \in [100, \infty]$ (or when $BF \in [-\infty, 1/100]$).

3.3 Results

Our results suggest that both visibility and distance are important for choosing whether and how to gesture when generating spatial referring expressions. Specifically, our Bayes Factor analysis suggests that while it is unlikely but uncertain whether distance directly informs referring gesture use (BF = 0.488, i.e., based on our data, it is about twice as likely that there is no main effect of distance on gesture use than that there is such a main effect), we can conclusively state that visibility directly informs referring gesture use (BF = 276,368, i.e., based on our data, it is over 250,000 times more likely that there is a main effect of visibility on gesture use than that there is no such effect). Moreover, our evidence allows us to conclusively state that distance and visibility interact to jointly inform gesture use (BF = 116, i.e., based on our data, it is over 100 times more likely that distance and visibility interact to jointly inform gesture use than it is that they do not).

Specifically, we observed that speakers were far more likely to use more precise gestures when their target was visible than when it was not (with 87.6% of the 97 visible referent descriptions using more precise gestures vs only 13.4% of the 157 non-visible referent descriptions using more precise gestures), and that when target referents were not visible, speakers became increasingly less likely to use more precise gestures and more likely to use more abstract gestures as their targets grew increasingly far away (with, for example, 22.2% of descriptions using more precise gestures and 51.9% of descriptions using more abstract gestures for referents at a distance of 12 feet, vs 7.1% of descriptions using more precise gestures and 71.4% of descriptions using more abstract gestures for referents at a distance of 270 feet).

The results of this experiment suggest a clear, simple policy for robot gesture design. While distance did play a factor in humans' choice of gestures, this effect was only in the relative frequency of more precise gestures as a minority class in those instances where most people chose to use a more abstract gesture due to referent non-visibility. As such, at least within environments like those examined where all objects of at least moderate distance are also occluded, robots may simply use more precise gestures for visible objects, and more abstract gestures for non-visible objects. To test the actual efficacy of such a policy, a second experiment was designed and conducted.

4 EXPERIMENT TWO

In our second experiment we seek to answer our second key research question: **(RQ2)** How do human-like referring gestures designed for large-scale interaction contexts (i.e., modeled on the more precise and more abstract gestures observed in Experiment One) objectively perform, and how are they subjectively perceived?

Specifically, we aimed to test four key hypotheses:

Hypothesis 1 (H1) — Abstract Pointing gestures will be objectively **more effective** in referring to non-visible objects; Precise Deictic gestures will be objectively more effective in referring to visible objects.

Hypothesis 2 (H2) — Abstract Pointing gestures will be perceived as **more human-like** when referring to non-visible objects; Precise Deictic gestures will be perceived as more humanlike when referring to visible objects.

Hypothesis 3 (H3) — Abstract Pointing gestures will be perceived as **more natural** when referring to non-visible objects; Precise Deictic gestures will be perceived as more natural when referring to visible objects.

Hypothesis 4 (H4) — Abstract Pointing gestures will be perceived as **more understandable** when used to refer to non-visible objects; Precise Deictic gestures will be perceived as more understandable when referring to visible objects

4.1 Method

4.1.1 Experimental Design. To test our hypotheses, we conducted a human-subjects study with two within-subject factors (Gesture Type and Referent Visibility) that also controlled for a three-way nuisance factor (Referent Direction), yielding a $2 \times 2 \times 3$ within-subjects Latin Square design.

The two *Gesture Type* conditions involved gestures of two different types: Precise Deictic (Fig. 3) and Abstract Pointing (Fig. 4). The two *Referent Visibility* conditions involved gestures toward objects that either were or were not visible to the user and robot. To control for effects of perspective, the three *Referent Direction* conditions involved gestures delivered towards objects in different directions with respect to the robot.

These category combinations were explored in a task environment containing six different objects, three of which were visible, and three of which were non-visible, organized into pairs of objects at nearly identical trajectories from the robot: two to the robot's left (one within the room and one outside the room), two to the robot's right, and two behind the robot.

Within this environment, the robot could thus point in one of three directions (ostensibly to one of six objects) using two different gesture types. To counterbalance participants' exposure to these six possible gestures, we designed a 6×6 balanced Latin Square of observable gestures. This produced a six-line table of condition sequences to which participants were randomly assigned. Meanwhile, the referent visibility factor was counterbalanced within-subjects using a repeated measure described later on.

4.1.2 Materials and Apparatus. To allow for more fine-grained control over our experimental environment and allow for the use of gestures that our physical robot platforms were not capable of (i.e., due to Pepper's lack of individually articulable fingers), our experiment was conducted in Virtual Reality (VR). Previous work suggests that physical and virtual robot gestures are perceived nearly identically [25], suggesting high potential for generalizability from virtuality to live interactions for the research questions we examined. For each of the six gestures described above (in one of the three directions using one of the two gesture types), we recorded a 4K 360°video showing a Softbank Pepper robot with fully articulable hands referring to a condition-determined sequence of six objects.

In these videos, Pepper performed Precise Deictic gestures by extending a straightened arm with its index finger pointed toward its target referent. While performing this gesture, Pepper's head turned to the object, not turning back until the gesture was complete. In contrast, Pepper performed Abstract Pointing gestures by extending an arm bent at the elbow, with an open palm oriented face up in the direction of the target referent. While performing this gesture, Pepper's head briefly turned in the direction of the target object before immediately turning back towards the participant.

To show the pre-recorded 360° videos of these gestures to participants, we used a Meta Quest 2 HMD: A commercial-grade VR headset that has an 1832×1920 LCD display per eye.

To facilitate replicability and reproducibility, all experiment materials, including videos with the Blender rendering file, questionnaires, Latin Square table, and data analysis, are available on Open Science Framework (OSF) at https://osf.io/nk4c7/.

4.1.3 Procedure. After providing informed consent, demographic information, and being instrumented with a VR headset, participants watched three tutorial videos on how to comfortably wear the headset, how to use its controller, and how to complete surveys within the VR headset. Experimenters proactively helped participants when needed and answered any clarification questions.

Each participant was then assigned to one of the six condition sequences, and watched the series of six 360° VR videos determined by that condition sequence. Before each video, participants were shown the map depicted in Fig. 5, and asked to familiarize themselves with it. Then, as defined by the experimental design, the participant watched one of the six videos determined by their assigned condition. Finally, participants were shown the map depicted in Fig. 5 again, and were asked which of the objects in the scene they believed the robot was referring to.

After viewing all six videos, participants were told to imagine that the robot had actually been referring to objects inside (or outside, for 50% of participants) the room. They were then asked to rewatch all six videos under this presumption, and after rewatching each video, were asked to evaluate the robot's gesture on the basis of how humanlike, natural, and understandable it was to them as a gesture towards the relevant object inside (or outside) the room. Finally, after rewatching all six videos, participants were told to re-imagine that in fact the robot had been referring to objects on the opposite side of the wall than they were previously told. They were then asked to rewatch all six videos for a third time under this opposite presumption, and re-rate the robot's gestures.

All experimenters followed an oral script to ensure consistency in experiment instructions. It took 40.5 minutes on average for participants to finish the whole study.

4.1.4 Measures. To test our four hypotheses, four measures were used as noted above, to separately assess effectiveness, humanlikeness, naturalness, and understandability.

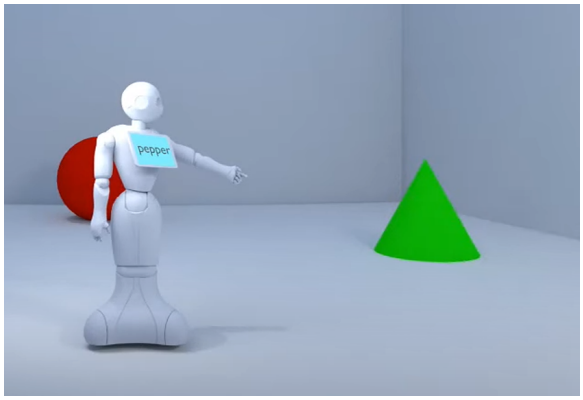

Figure 3: Pepper using a Precise Deictic gesture to refer to a visible object (the green cone). The robot stayed turned at the object with sustained gaze.

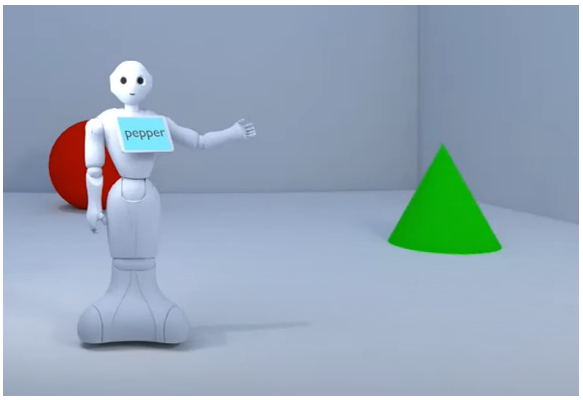

Figure 4: Pepper using an Abstract Pointing gesture to refer to one of the three non-visible objects (the blue cube beyond the rightmost wall in Fig. 5). The robot briefly glanced at the object and turned back.

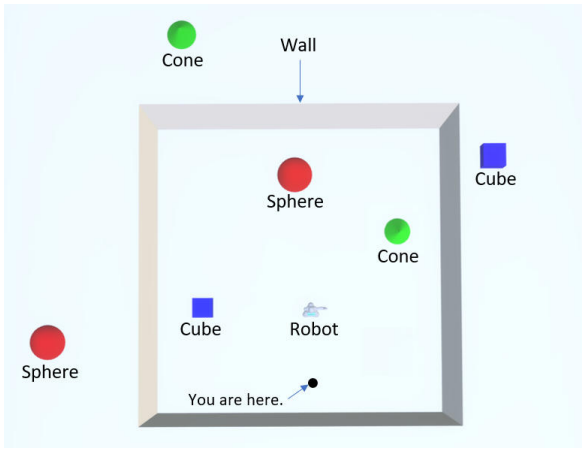

Figure 5: A top-down of the VR task environment. The Pepper robot was placed in the bottom center of the room, in front of the user. The environment contains three visible objects inside the wall and three non-visible objects outside.

Effectiveness was measured by assessing whether participants' guesses at the intended target of each robot gesture would have been correct under a policy in which Abstract Pointing gestures are used to refer only to non-visible objects, and Precise Deictic gestures are used to refer only to visible objects. This thus represents not the participant's effectiveness, but rather the effectiveness that that hypothetical gesture policy would have facilitated.

Humanlikeness was measured using the 5-point Godspeed Anthropomorphism Scale [5].

Naturalness was measured using a 5-point Likert item asking participants how natural the robot's gesture was.

Understandability was measured using a 5-point Likert item asking participants how understandable the robot's gesture was.

4.1.5 Analysis. Our data were analyzed using Bayesian Repeated Measures Analyses of Variance (RM-ANOVAs) with Bayes Factors calculated across matched models [39], using JASP 0.18.

4.1.6 Participants. 34 participants were recruited from Colorado School of Mines. Of these, 13 (38.23%) identified as women, 19 (55.88%) identified as men, and 2 did not wish to disclose their gender. For racial identity, 18 identified as White (52.94%), 8 as Asian (23.53%), 3 as belonging to more than one racial group (8.824%), 1 as Latino (2.941%), and 4 chose not to disclose. Participant ages ranged from 18 to 41 (M=22.73, SD=5.69). 20 reported familiarity with robots, 6 were neutral, and 8 reported being unfamiliar with robots. 13 reported being familiar with virtual reality, 4 reported neutral experience, and 17 were unfamiliar with virtual reality. Each was given a $15 Amazon gift card for their participation.

4.2 Results

4.2.1 Effectiveness. A repeated measures Analysis of Variance (RM-ANOVA) revealed extreme evidence for an effect of Gesture ($BF_{10} = 1.190 \times 10^5$). As shown in Fig. 6, when the robot used a precise deictic gesture, around 80% of participants ($M = 79.4$, $SD = 18.4$) believed the robot was talking about something that was visible (whereas around 20% of participants thought it was talking about something non-visible). Meanwhile, when the robot used an Abstract Pointing gesture, only around 40% of participants ($M = 43.1$, $SD = 33.4$) believed the robot was talking about something non-visible (while around 60% of participants thought it was talking about something visible). This suggests that using Precise Deictic gesture is indeed the most effective way to refer to something visible (as 80%>60%), and that using Abstract Pointing gesture is indeed the most effective way to refer to something non-visible (as 40%>20%), but that on the other hand, using gesture alone is unlikely to be a strong enough signal to pick out a non-visible object due to a (very reasonable) interpretation bias toward visible objects. These results thus supports **H1** (while highlighting that an expectation of relying on gesture alone is, perhaps, unrealistic).

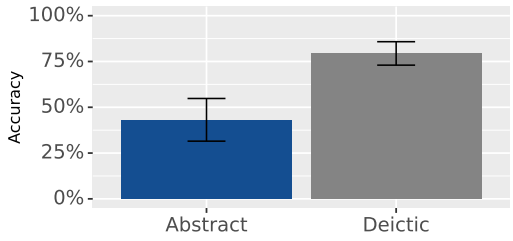

Figure 6: Objective effectiveness. Error bars in this and all later charts show a 95% credible interval. Results show that while Abstract Gestures are not readily interpreted on their own, using Precise Deictic gestures to refer to visible objects and using Abstract Pointing gestures to refer to non-visible objects is the best policy given our data.

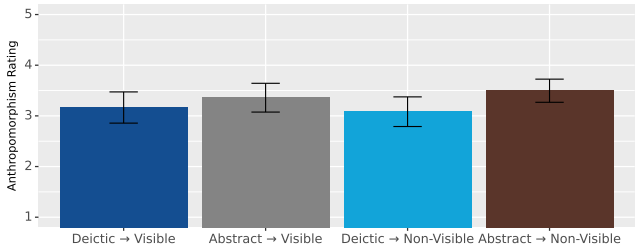

Figure 7: Anthropomorphism. Results show a difference between Precise Deictic and Abstract Pointing gestures towards non-visible objects.

4.2.2 Anthropomorphism. A two-way repeated measures Analysis of Variance (RM-ANOVA) revealed strong evidence for an effect of Gesture ($BF_{10} = 22.199$). As shown in Fig. 7, participants viewed robots that used Abstract Pointing gestures as more humanlike ($M = 3.427$, $SD = 0.735$) than robots that used Precise Deictic gestures ($M = 3.124$, $SD = 0.861$).

This RM-ANOVA also revealed moderate evidence against an effect of Referent Visibility ($BF_{10} = 0.251$).

Finally, this RM-ANOVA revealed moderate evidence in favor of an interaction between Gesture and Referent Visibility ($BF_{10} = 5.059$). Specifically, post-hoc Bayesian t-tests revealed that the difference in ascriptions of humanlikeness for Abstract Pointing vs Precise Deictic gestures was very strong for gestures to non-visible objects ($M_A = 3.496, SD_A = 0.655$ vs. $M_D = 3.082, SD_D = 0.838$; $BF_{10} = 80.733$) but that there was actually only anecdotal evidence *against* such a difference for gestures to visible objects ($M_A = 3.359, SD_A = 0.814$ vs. $M_D = 3.165, SD_D = 0.883$; $BF_{10} = 0.693$). These results partially support **H2**: Abstract Pointing gestures were perceived as more humanlike when referring to non-visible objects, but no such benefit was seen for Precise Deictic gestures when referring to visible objects.

4.2.3 Naturalness. A two-way repeated measures Analysis of Variance (RM-ANOVA) revealed anecdotal evidence for an effect of Gesture ($BF_{10} = 2.253$). As shown in Fig. 8, there was not enough

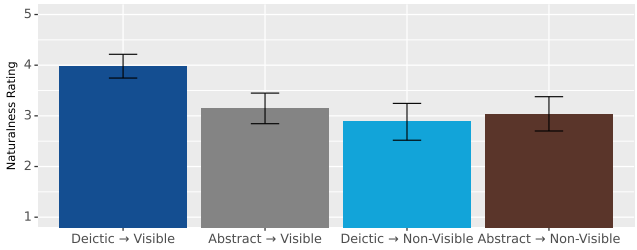

Figure 8: Naturalness. a difference between Precise Deictic and Abstract Gestures towards visible objects.

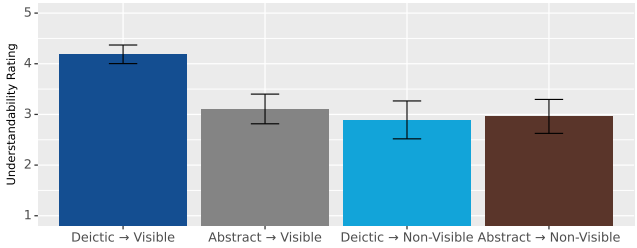

Figure 9: Understandability. Results show a difference between Precise Deictic and Abstract Pointing gestures towards visible objects.

evidence to conclusively support or rule out an effect, but the evidence tentatively suggests that if there were one, it would be that participants viewed robots that used Precise Deictic gestures as more natural ($M = 3.431$, $SD = 0.856$) than robots that used Abstract Pointing gestures ($M = 3.093$, $SD = 0.918$).

This RM-ANOVA also revealed very strong evidence for an effect of Referent Visibility ($BF_{10} = 54.782$). Specifically, participants viewed robots as more natural when they were referring to visible objects ($M = 3.564, SD = 0.768$) than when they were referring to non-visible objects ($M = 2.961, SD = 1.006$).

Finally, this RM-ANOVA revealed extreme evidence in favor of an interaction between Gesture and Referent Visibility ($BF_{10} = 1.949 \times 10^5$). Post-hoc Bayesian t-tests revealed that the difference in ascriptions of naturalness for Abstract Pointing vs Precise Deictic gestures was extreme for gestures to visible objects ($M_A = 3.147, SD_A = 0.865$ vs. $M_D = 3.980, SD_D = 0.671$; $BF_{10} = 6265.049$) but that there was moderate evidence *against* such a difference for gestures to non-visible objects ($M_A = 3.039, SD_A = 0.970$ vs. $M_D = 3.039, SD_D = 0.970$; $BF_{10} = 0.260$). These results partially support **H3**: Precise Deictic gestures were perceived as more natural when referring to visible objects, but no such benefit was seen for Abstract Pointing gestures when referring to non-visible objects.

4.2.4 Understandability. A two-way repeated measures Analysis of Variance (RM-ANOVA) revealed strong evidence for an effect of Gesture ($BF_{10} = 19.122$). As shown in Fig. 9, participants viewed robots that used Precise Deictic gestures as more understandable ($M = 3.359$, $SD = 0.9$) than robots that used Abstract Pointing gestures ($M = 3.035$, $SD = 0.8$).

This RM-ANOVA also revealed extreme evidence for an effect of Referent Visibility ($BF_{10} = 434.379$). Specifically, participants viewed robots as more understandable when they were referring to visible objects ($M = 3.647, SD = 0.684$) than when they were referring to non-visible objects ($M = 2.927, SD = 1.016$).

Finally, this RM-ANOVA revealed extreme evidence in favor of an interaction between Gesture and Referent Visibility ($BF_{10} = 6.415 \times 10^7$). Specifically, post-hoc Bayesian t-tests revealed that the difference in ascriptions of understandability for Abstract Pointing vs Precise Deictic gestures was extreme for gestures to visible objects ($M_A = 3.108, SD_A = 0.840$ vs. $M_D = 4.186, SD_D = 0.527$; $BF_{10} = 6.159 \times 10^4$) but that there was actually moderate evidence *against* such a difference for gestures to non-visible objects ($M_A = 2.961, SD_A = 0.960$ vs. $M_D = 2.892, SD_D = 1.072$; $BF_{10} = 0.194$). These results partially support **H4**: Precise Deictic gestures were perceived as more understandable when referring to visible objects, but no such benefit was seen for Abstract Pointing gestures referring to non-visible objects.

5 DISCUSSION

5.1 Abstract Pointing (when Multimodal) is More Effective for Non-Visible Objects

Our first hypothesis was about effectiveness: that Abstract Pointing gestures would be more effective in referring to non-visible objects, and that Precise Deictic gestures would be more effective in referring to visible objects. Our results support both facets of this hypothesis. Most participants in the Precise Deictic condition (80% vs 60% compared to Abstract Pointing) were able to infer that a visible object was being referenced. For non-visible objects, Abstract Pointing gestures were more effective but the accuracy was only 40% vs 20% compared to Precise Deictic gestures. This shows the promise of using Abstract Pointing gestures to refer to non-visible objects. The relatively low accuracy of these gestures is likely due only to the complete reliance on gesture in this experiment: trying to infer the target of potentially non-visible objects from non-verbal cues alone is extremely challenging, both due to the ambiguity of non-verbal communication and due to the human interpretation bias towards visible objects. This finding aligns with work showing the need for verbal explanation by Han et al. [24]. In contrast, pairing abstract gestures with spoken language would likely lead to acceptable accuracy. Future work should be performed to confirm this. Thus, we propose **Design Guideline 1: Abstract Pointing gestures can be used to help identify non-visible referents, but should always be accompanied with information conveyed through other communication modalities.**

5.2 Abstract Pointing Increases Anthropomorphism

Our second hypothesis was that Abstract Pointing gestures to non-visible objects would appear more human-like and that Precise Deictic gestures to visible objects would appear more human-like. Our results only partially support this hypothesis. When made towards non-visible objects, Abstract Pointing gestures were more humanlike than Precise Deictic gestures; but for gestures towards visible objects, Precise Deictic gestures were no more humanlike (and if anything, were less humanlike than Abstract Pointing Gestures). Overall these results suggest that using abstract pointing gestures towards non-visible objects may be an effective strategy if one wishes to invoke attributions of human characteristics, activate familiar interactions, and encourage willingness to interact and accept robot behaviors [36]. And, conversely, the use of such gestures should be avoided if one is concerned about over-anthropomorphization of a robot. As such, we propose **Design Guideline 2: The use of Abstract pointing gestures to non-visible objects should be informed in part by designers' desire to encourage or discourage anthropomorphism**.

5.3 Precise Dectic Gestures to Visible Objects are More Natural and Understandable

Our third and fourth hypothesis were that Abstract Pointing gestures would appear more natural (H3) and understandable (H4) when referring to non-visible objects and that Precise Deictic gestures would appear more natural (H3) and understandable (H4) when referring to visible objects. Our results showed that deictic gestures were more natural and understandable when referring to visible objects, but that there was no difference between the gestures when referring to non-visible objects. As such, we propose **Design Guideline 3: When robots refer to visible objects, they should use Precise Deictic gestures, i.e., with direct, and sustained use of both deictic gaze and deictic pointing**.

5.4 Limitations and Future Work

As discussed in Sec 5.1, future work should further investigate the efficacy of Abstract Pointing in the context of multimodal referring utterances. In addition, future work should address key limitations of this experiment. First, while our use of a virtual environment helped to provide us with enhanced experimental control, environmental control, and overcome the inherent limitations of today's robotic hardware, future work will ultimately be needed with physical robots situated in real physical environments. Second, future work should explore even larger-scale environments and referents that are either farther away yet still visible, or that are much farther away not possibly visible. Finally, future work should explore the other types of referential gestures from Higger et al. [27]'s taxonomy, and how these gestures might be used based on factors other than visibility or distance, such as known-ness or uncertainty.

6 CONCLUSIONS

In this work, we identified key factors in the human use of different types of referring gestures, including Precise Deictic and Abstract Pointing gestures. We then investigated robots' use of Precise Deictic and Abstract Pointing gestures in reference to both visible and non-visible objects. Our results show that while the benefits of each gesture type are reflected through different metrics, there is an overall benefit to using Precise Deictic gestures when referencing visible objects, and Abstract Pointing gestures (accompanied by informative verbal cues) when referring to non-visible objects.

ACKNOWLEDGMENTS

This work was supported in part by NSF grant IIS-1909864 and by ONR grant N00014-21-1-2418.

REFERENCES

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

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

[3] Martha W Alibali. 2005. Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial cognition and computation* 5, 4 (2005), 307–331.

[4] Gary L Allen. 2003. Gestures accompanying verbal route directions: Do they point to a new avenue for examining spatial representations? *Spatial cognition and computation* 3, 4 (2003), 259–268.

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

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

[7] Cynthia Breazeal, Cory Kidd, Andrea Lockerd Thomaz, et al. 2005. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *Proceedings of the i (IROS)*.

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

[9] Paul Bremner, Anthony G Pipe, Mike Fraser, Sriram Subramanian, and Chris Melhuish. 2009. Beat gesture generation rules for human-robot interaction. In *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 1029–1034.

[10] Andrew G Brooks and Cynthia Breazeal. 2006. Working with robots and objects: Revisiting deictic reference for achieving spatial common ground. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.

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

[12] Elizabeth Cha, Yunkyung Kim, Terrence Fong, and Maja J Mataric. 2018. A Survey of Nonverbal Signaling Methods for Non-Humanoid Robots. *Found. Trend. Rob.* (2018).

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

[14] Eve Clark and C Sengul. 1978. Strategies in the acquisition of deixis. *J. Child Lang.* (1978).

[15] Antonella De Angeli, Walter Gerbino, Giulia Cassano, and Daniela Petrelli. 1998. Visual display, pointing, and natural language: the power of multimodal interaction. In *AVI*.

[16] Nick J Enfield, Sotaro Kita, and Jan Peter De Ruiter. 2007. Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics* 39, 10 (2007), 1722–1741.

[17] Rui Fang, Malcolm Doering, and Joyce Y Chai. 2015. Embodied collaborative referring expression generation in situated human-robot interaction. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. 271–278.

[18] Albert Gatt and Patrizia Paggio. 2014. Learning when to point: A data-driven approach. In *Proc. Int'l Conf. on Computational Linguistics (COLING)*. 2007–2017.

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

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

[21] Susan Goldin-Meadow and Martha Wagner Alibali. 2013. Gesture's role in speaking, learning, and creating language. *Annual review of psychology* 64 (2013), 257–283.

[22] Marianne Gullberg. 1996. Deictic gesture and strategy in second language narrative. In *Workshop on the Integration of Gesture in Language and Speech*.

[23] Jared Hamilton, Thao Phung, Nhan Tran, and Tom Williams. 2021. What's The Point? Tradeoffs Between Effectiveness and Social Perception When Using Mixed Reality to Enhance Gesturally Limited Robots. In *Proceedings of the 16th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.

[24] Zhao Han, Elizabeth Phillips, and Holly A Yanco. 2021. The need for verbal robot explanations and how people would like a robot to explain itself. *ACM Transactions on Human-Robot Interaction (THRI)* 10, 4 (2021), 1–42.

[25] Zhao Han, Yifei Zhu, Albert Phan, Fernando Sandoval Garza, Amia Castro, and Tom Williams. 2023. Crossing Reality: Comparing Physical and Virtual Robot Deixis. In *ACM/IEEE International Conference on Human-Robot Interaction*.

[26] Simon Harrison. 2011. The creation and implementation of a gesture code for factory communication. *Proc. Int. Conf. on Gesture in Speech and Interaction* (2011).

[27] Mark Higger, Polina Rygina, Logan Daigler, Lara Ferreira Bezerra, Zhao Han, and Tom Williams. 2023. Toward Open-World Human-Robot Interaction: What Types of Gestures Are Used in Task-Based Open-World Referential Communication?. In *The 27th Workshop on the Semantics and Pragmatics of Dialogue (SemDial)*.

[28] Rachel M Holladay, Anca D Dragan, and Siddhartha S Srinivasa. 2014. Legible robot pointing. In *The 23rd IEEE International Symposium on robot and human interactive communication*. IEEE, 217–223.

[29] Rachel M Holladay and Siddhartha S Srinivasa. 2016. RoGuE: Robot Gesture Engine. In *Proceedings of the AAAI Spring Symposium Series*.

[30] Chien-Ming Huang and Bilge Mutlu. 2013. Modeling and Evaluating Narrative Gestures for Humanlike Robots.. In *Robotics: Science and Systems*. 57–64.

[31] Chien-Ming Huang and Bilge Mutlu. 2014. Learning-based modeling of multimodal behaviors for humanlike robots. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 57–64.

[32] Jana M Iverson and Susan Goldin-Meadow. 1998. Why people gesture when they speak. *Nature* 396, 6708 (1998), 228–228.

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

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

[35] Adam Kendon. 2000. Language and gesture: Unity or duality. *Language and gesture 2* (2000).

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

[37] Michael D Lee and Eric-Jan Wagenmakers. 2014. *Bayesian cognitive modeling: A practical course*. Cambridge university press.

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

[39] S. Mathôt. 2017. Bayes like a Baws: Interpreting Bayesian repeated measures in JASP [Blog Post].

[40] David McNeill. 1992. *Hand and mind: What gestures reveal about thought*. University of Chicago press.

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

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

[43] Paul Piwek. 2009. Salience in the generation of multimodal referring acts. In *Proceedings Int'l Conf. on Multimodal interfaces*. ACM, 207–210.

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

[45] Maha Salem, Stefan Kopp, et al. 2010. Towards an integrated model of speech and gesture production for multi-modal robot behavior. In *Proc. of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*.

[46] Maha Salem, Stefan Kopp, Ipke Wachsmuth, Katharina Rohlfing, and Frank Joublin. 2012. Generation and evaluation of communicative robot gesture. *Int. Jour. Social Robotics* (2012).

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

[48] Mandana Seyfeddinipur and Sotaro Kita. 2001. Gestures and self-monitoring in speech production. In *Annual Meeting of the Berkeley Linguistics Society*, Vol. 27. 457–464.

[49] Nhan Tran, Trevor Grant, Thao Phung, Leanne Hirshfield, Christopher Wickens, and Tom Williams. 2023. Now Look Here! ↓ Mixed Reality Improves Robot Communication Without Cognitive Overload. In *International Conference on Virtual, Augmented, and Mixed Reality (VAMR), held as part of the International Conference on Human-Computer Interaction (HCI)*.

[50] Ielka Francisca Van Der Sluis. 2005. *Multimodal Reference, Studies in Automatic Generation of Multimodal Referring Expressions*. Ph. D. Dissertation. University of Tilburg.

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

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

[53] William Dwight Whitney. 1875. *The life and growth of language*. Vol. 16. HS King.

[54] Tom Williams, Saurav Acharya, Stephanie Schreitter, and Matthias Scheutz. 2016. Situated open world reference resolution for human-robot dialogue. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 311–318.

[55] Tom Williams, Matthew Bussing, Sebastian Cabrol, Elizabeth Boyle, and Nhan Tran. 2019. Mixed Reality Deictic Gesture for Multi-Modal Robot Communication. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.

[56] Tom Williams and Matthias Scheutz. 2015. POWER: A domain-independent algorithm for probabilistic, open-world entity resolution. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 1230–1235.

[57] Tom Williams and Ruchen Wen. 2021. Human Capabilities as Guiding Lights for the Field of AI-HRI: Insights from Engineering Education. In *AAAI Fall Symposium on Artificial Intelligence for Human-Robot Interaction (AI-HRI)*.